

In pursuit of comprehension

Nature is injecting increased effort into the readability of its papers. This is a task in which authors can help by trying out their texts on colleagues before submission.

To have "intimate relations with the little band... of nature's servants in every civilized region of our planet" was one of this publication's expressed intentions at its outset, in November 1869. But intimacy can breed withering exclusiveness, rather than stimulating inclusiveness, if it expresses itself in private language and with hidden rules. Such an approach is at odds with one of *Nature's* original prime ambitions: "to place before the general public the results of scientific work..." and downright harmful to another: "to aid Scientific Men themselves by giving early information of all advances made in any branch of natural knowledge throughout the world and by affording them an opportunity of discussing the various scientific questions which arise from time to time..."

Exclusiveness is all too frequent nowadays. Like other professions, science thrives on language that becomes shared by practitioners only by dint of their participation and which is therefore often, in effect, private. Similarly, science flourishes on hard-won knowledge of phenomena and principles that rapidly becomes taken for granted by cognoscenti. To the extent that this knowledge is consequently implicit in publications, it is hidden from outsiders. Worse still for readers with wide ranging interests, the range of esoteric (and sometimes downright whimsical) terminology is growing rapidly.

Transmitting such breadth without compromising scientific impact is still one of *Nature's* goals, and is, thankfully, one to which an increasing number of readers are willing to subscribe. Regrettably, the general public may benefit from our news, opinions and correspondence, but will feel well and truly excluded from the original science. But nowadays the papers — the Articles and Letters — too often prove daunting even to a broad scientific readership. Is that situation inevitably going to get even worse?

Nature believes, on the contrary, that things can get at least a little better. On the one hand, recent internal exercises by editorial staff exploring the potential comprehensibility of Letters indicate that it would be a hard task indeed to make the introductions of, for example, all molecular biological papers easily understandable by, say, geologists. On the other hand, those same exercises suggest that the introductions of many life-sciences Letters could be more readily accessible to most biologists, that an analogous task can be tackled in the physical sciences, and, most encouragingly, that there are still a significant number of occasions when introductions can be presented so that all readers have a hope of quickly grasping a paper's principal conclusion.

More editorial attention is accordingly being placed on the readability of the introductory summaries of our Letters. Feedback suggests that those efforts are bearing fruit, though there is more to be done. But the people who can do most to help are our authors. Acceptance of a paper in any journal, after all, represents not only a reward but also an opportunity to maximize impact — and all the more so when the journal is as widely read as *Nature*.

To that end, it is salutary for authors writing the introduction to a paper to consider at what stage in their education they first came across terms they use. (They might then be less shocked on realizing that, for example, some respectable biologists do not know what "anisotropic" means, more than a few physicists do not know what a phylum is, and that there are plenty of both who have forgotten the meaning of "epimerization".) In considering their

work's readability by scientists in other disciplines, authors might take the trouble to refer briefly to classic textbooks to remind themselves what comprehension cannot be taken for granted. More directly, they can try their own texts out on appropriately remote colleagues.

They should at least respond positively when *Nature's* editors prod them to remove the *n*th unexplained acronym in their first paragraph. And they can reflect on what anyone in the book trade knows well: no science publisher ever lost sales by underestimating the knowledge of specialists outside their specialities. □

Winning is not everything

Germany is to use competitions to boost university-industry collaboration. The idea has merits — but also pitfalls.

COMPETITIONS of all types usually have two functions: not only do they reward outstanding achievement but — at least in principle — they can also stimulate improved performance in those who take part. Germany's research minister Jürgen Rüttgers seems to have been taken with this idea as a way of supporting closer collaboration between academic scientists and industry. Last month, when he announced the results of the BioRegio competition, designed to challenge regions to develop ideas for boosting biotechnology, even losers admitted that they had benefited from the experience of taking part.

Entering the competition had required considerable efforts in preparing detailed proposals for collaboration. The main prize turns out to have been the prestige of winning. But all the entries still exist — as do the contacts between academics and industrialists that developed during their preparation.

Encouraged by the apparent success of his strategy, Rüttgers has introduced four more competitions (see page 500). This time, his challenge to both researchers and industry includes the task of defining *Leitprojekte* — research areas of strategic economic and social importance. The idea is not unique to Germany; Britain recently awarded financial support (and the accompanying prestige) to collaborative projects selected as winners in its Foresight Challenge competition, based on priorities identified through the Technology Foresight programme.

Competitions can certainly help to overcome the inertial forces that have tended to keep industry and academic institutions apart. But Rüttgers must be careful not to take the idea too far. For one thing, prizes are to be taken from existing — and shrinking — ministry budgets. If too much public money is set aside for the market-orientated research programmes that the competitions are designed to inspire, other funding could suffer badly.

Second, Rüttgers should remember that part of the attraction of the BioRegio was its novelty — and thus the relative inexperience of participants in such activities. Once they become aware that entering competition requires considerable effort, with the prospects of little financial reward, enthusiasm may evaporate. More substantial prizes may then be needed to maintain momentum. Could German science afford it? □



Missing dinosaur skulls raise new fears of smuggling in Moscow

Munich. Palaeontologists are growing increasingly concerned about the disappearance of at least five valuable dinosaur skulls and other fossils from the Museum of the Palaeontological Institute in Moscow.

Scientists are worried that the skulls may have been smuggled abroad and sold on the black market. But the reluctance of senior officials to discuss the disappearances — even with the police — appears to be frustrating attempts to solve the problem. The institute, part of the Russian Academy of Sciences, houses one of the world's most important fossil collections.

Following publicity over a wave of thefts in 1992, officials from the institute agreed to participate in an international working group set up in 1994 to help locate and return stolen Russian fossils (see *Nature* 371, 729; 1994). The group is led by David Unwin, of Bristol University's Department of Geology, and Michael Shishkin, of the Palaeontological Institute in Moscow.

The officials also agreed to improve security at the institute to prevent further disappearances, and to terminate all contact with commercial organizations selling fossils. Smaller invertebrate and vertebrate fossils have nevertheless regularly continued to disappear — to the frustration of scientists at the institute, who have been told not to speak publicly about any disappearances they notice.

The recent vertebrate disappearances have only increased concern. They include skulls of two *Protoceratops* and a *Breviceratops kozlowskii*, and the upper and lower jaw of the species *Tarbosaurus efremovi*, both jaws being more than half a metre long. The five fossils disappeared from locked basement stores sometime before August, when the absence was discovered.

The *Tarbosaurus* lower jaw is a particularly significant loss to science. It is the world's principal reference specimen, and is central to attempts to answer fundamental questions about the dinosaur, such as whether it is the same species as *Tyrannosaurus rex*.

The value of these five dinosaur fossils is estimated to be around US\$11,000. Although the apparent theft of at least one collection of invertebrate fossils has been reported to the police, the more serious disappearance of these vertebrate fossils has not — to the distress of institute scientists.

Alexei Rozanov, the director of the institute, declines to comment on the apparent reluctance to report the disappearances. But Igor Novikov, his deputy director, justifies it on the grounds that “we cannot expect much help in such cases from the police, either Russian or Interpol”.

Nevertheless, after an enquiry last month from the western members of the international working group for the return of stolen

generous daily allowance in foreign currency, as well as the opportunity to visit foreign palaeontological museums and institutes. These inducements, as well as the fear of losing their jobs, is said to have encouraged staff to say nothing about missing fossils. But it has led to discontent among some of these scientists within the institute.

Novikov, speaking for the institute, says that its directors are cooperating with various state bodies, including customs, “to put the fossil trade under the control of experts”. He says that stronger locks have been put on laboratory doors, and access to material has been restricted.

He argues that the concerns of western scientists reflect prejudice against their colleagues from former Communist countries, and suggests that they should focus less on his institute and more on the burgeoning black market in fossils in the West, which drives demand for stolen material.

Novikov also accuses some westerners of exploiting the institute's collection. He says Bernard Bredow, owner of a small museum in Bavaria called Mammuthem, kept material well beyond an agreed loan period, and made casts of fossils without permission.

Bredow replies that delays in returning some specimens used in an Austrian exhibition were caused by problems at customs in Austria and Moscow. He says that he signed a legal agreement in 1991 with the institute's former director, Leonid Tatarinov, giving him full responsibility for the institute's activities in Germany and the right to make casts of fossils.

“This contract is still valid,” says Bredow. But he says he has avoided contact with the institute since 1992, after concern about dealers to whom he had been introduced by the institute, and who snatched fossils that he had originally bought from them while he on display at a trade fair in Munich.

Bredow says that the argument with the dealers arose because he had not paid all of the three separate bills he received for the same items. Police were called in, and the fossils — which were not part of the museum's collection — were retrieved.

He says that he is keen to help the institute, but that this had become impossible because of a lack of “transparency” that made it difficult to know what was going on. But Bredow says that he will do what he can to help locate the missing fossils.

Alison Abbott

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Missing link: skull of *Tarbosaurus efremovi*, part of which has recently disappeared from storage.

Russian fossils, and apparently concerned by rumours of the theft, Novikov confirmed the loss of the five dinosaur fossils, and provided descriptions and internal catalogue numbers. But, despite repeated requests from the working group, Novikov has not provided illustrations needed for an international search for the fossils.

Novikov did not confirm the disappearance of historically important mammoth tusks and a large number of ammonites, which some scientists say are missing.

The international working group is planning to publish details of the dinosaur fossils in palaeontological journals, as well as on the Internet, in the hope that they may be located and returned. The prospects are not good; only one of about 50 fossils stolen two years ago has so far been returned.

Since the fall of Communism, the Palaeontological Institute, like all publicly funded research institutes in Russia, has been in severe financial difficulties and has had to find new ways to raise money. In the early 1990s, staff members set up businesses for the sale of fossils using the institute's address. But this practice has now been stopped, says Novikov.

Stolen fossils could be leaving the country illegally through St Petersburg port or Moscow airport. In theory they could also be smuggled out of the institute in crates used to send specimens to exhibitions abroad.

The institute raises about US\$30,000 a month by mounting exhibitions around the world. Some personnel are allowed to accompany the exhibitions and receive a

US drops patent claim to Hagahai cell line

San Francisco. The US National Institutes of Health (NIH) has dropped a controversial claim to the patent for a cell line derived from a Hagahai man of Papua New Guinea which had been criticized by activist groups that oppose patents on living organisms — including human tissue.

The patent, issued to the Department of Health and Human Services in March 1995, was the first to be based on a human cell line taken from an indigenous population, and was based on the fact that the cell line contained a unique variant of the human t-lymphotropic virus. The notice of the disclaimer was due to be published on Tuesday (10 December) in the *Official Gazette* of the US Patent and Trademark office.

NIH filed papers to forfeit its rights to the patent claims at the patent office in October. Barbara McGarey, deputy director of the NIH's Office of Technology Transfer, says that after initially filing patent applications on a wide variety of biological products, the NIH decided to limit such applications to products with significant commercial potential, and the Hagahai cell line patent therefore became unnecessary. "We have stopped filing patents on what are primarily research tools," she said.

The US government had shared invention on the patent with Carol Jenkins, principal research fellow at the Institute of Medical Research in Papua New Guinea.

Jenkins, who had agreed to hand over her 50 per cent of the royalties to the Hagahai people, eventually apparently decided that the difficulty and expense of setting up a trust outweighed any potential benefit from the patent, according to a source. The NIH, which had long made known its lack of interest in the patent, readily agreed.

Originally Jenkins, who is involved in AIDS intervention research in Papua New Guinea's central highlands, had agreed with the Hagahai people that patenting was the best approach to protecting their interests. But a patent would cost up to US\$17,000 to maintain over its 17-year lifespan, not

including the costs of maintaining a trust.

Both parties had been aware from the beginning that the likelihood of any commercial value from the HTLV-1 variant protected by the patent was small. But at the time, scientists involved with the research said they felt compelled to establish some precedent to protect the Hagahai's interest in their own tissue.

The Hagahai patent became the focus of international controversy, and Jenkins narrowly avoided deportation by the Papua New Guinea government, which later cleared her of any wrongdoing and concluded that her work had been conducted with the full consent of the Hagahai people (see *Nature* **380**, 374; 1996).

She stayed and agreed to work with government officials to develop a policy on appropriate conduct for research on indigenous populations. Eventually Jenkins, her institute and the Hagahai decided to disclaim the patent.

The Rural Advancement Foundation International, which had condemned the patent and HGDP for affronting human rights and treating human tissues as commodities, claimed victory following the NIH's patent disclaimer. "I hope this is the end of what is arguably the most offensive patent ever issued," says Alejandro Argumedo of the Indigenous People's Biodiversity Network in Canada. **Sally Lehrman**

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Planet Earth/John Downer

Hunting genes: the Hagahai patent attracted criticism but research had their consent.

Germany extends academic-industry competition plans

Munich. Germany's federal research ministry is to launch a series of competitions next month intended to reward joint teams of academic and industrial scientists for the best ideas for new products or processes.

The idea is part of an attempt by Jürgen Rüttgers, the research minister, to avoid funding only those research projects that fall under strict headings laid down by the ministry, and to let the research community itself identify leading projects (*Leitprojekte*).

Competitions will be launched in four areas: molecular medicine, new materials and miniaturization technologies, information technology in education and mobility in large cities. Winning projects will be funded for about five years.

In a first phase, teams will be invited to provide a description of their general ideas. About a dozen of these ideas will be selected by a jury to take part in the second phase — to be subsidized by the ministry — in which fuller applications describing specific projects must be submitted. The winner or winners (Rüttgers has not decided how many there will be) will probably be announced next autumn.

The idea of funding *Leitprojekte* comes from pressure from industry to establish

closer connections with state-funded research institutions. In a document published last July, the Bundesverband der Deutschen Industrie (BDI), the group representing German industry in its negotiations with government, recommended that researchers from the academic community and industry should agree with politicians on the strategic research areas of economic or social benefit to Germany, and on realistic research programmes within those areas.

According to the BDI, such research programmes, which it termed *Leitprojekte*, would be carried out jointly by industry and state-financed institutions. Rüttgers' idea of introducing *Leitprojekte* through competitions has been inspired by the success of his recent 'BioRegio' competition (see *Nature* **384**, 298; 1996), in which consortia of academic institutes and companies were encouraged to come up with ideas for developing biotechnology in their regions.

As in the BioRegio competition, Rüttgers has not specified the conditions of the *Leitprojekte* contests, or the nature of the prizes. But about DM100 million (US\$64.5 million) is expected to be made available out of existing ministry budgets, and the scheme could command up to 10 per cent of the ministry's

normal budget for industrial research. (Participants are also likely to be asked to contribute significantly to project costs.)

The scheme has the enthusiastic approval of Germany's industrial research community. Heinrich Höfer, head of the BDI's department of technology and innovation politics, says the idea is a "winning formula". Basic research organizations, such as the Max Planck Society and the Helmholtz Society, which represents Germany's 16 national research centres, are more circumspect.

Detlev Ganten, head of the Max Delbrück Centre for Molecular Medicine, a national research centre in Berlin, acknowledges the advantages of the competition idea. Although his region was not a winner of the BioRegio competition, he recognizes its leverage effect. "It was a good way to concentrate research effort and mobilize people, for very little money," he says.

But Ganten shares the concern of many scientists about Rüttgers' desire to extend the idea, particularly if it goes too far. With ever-decreasing industrial and government funding, he says, basic research could lose out if an increasing proportion of available money is directed toward strategic research goals.

Alison Abbott

NIH asked to budget for flat funding for management costs

Washington. Harold Varmus, director of the US National Institutes of Health (NIH), has been told by a leading congressman to assume that the agency's budget for research management and support will be held flat for the three fiscal years that begin in 1998. This support budget was cut by 7.5 per cent, to \$481 million, in 1996, a figure that was maintained in 1997, the current fiscal year that began on 1 October.

John Porter (Republican, Illinois), chairman of the subcommittee of the House appropriations committee that oversees NIH's \$12.7 billion budget, made known his plans for the support budget late last month in a letter to Varmus. He asked Varmus to conduct a comprehensive review of the NIH's administrative structure and costs by next June, and to produce a long-range research management plan assuming "level funding" — with no inflation adjustment — for the next three years.

Officials have argued that flat administrative budgets make it difficult to support the growing extramural activity stimulated by the agency's overall budget boosts in 1996 and 1997. But an aide to Porter points out that administrative budgets were held flat in 1997 across all the departments in the subcommittee's jurisdiction, including labor, health and human services and education. the same." **M. W.**

Gibbons 'has no plans to leave government'

Washington. Jack Gibbons, director of the White House Office of Science and Technology Policy (OSTP) and science adviser to President Bill Clinton, said last week he "has no plans to leave the government at this time", and has laid out an agenda for OSTP in the first half of 1997. But he did not commit himself to serving a full second term with the Clinton administration, saying he had yet to discuss it with the president.

Gibbons said that the office will continue to work on international scientific collaboration, engage in "constructive dialogue" with Congress on funding for the National Aeronautics and Space Administration (NASA) and the future structure of the Department of Energy, and seek to develop an "innovative partnership" with state governments.

The office plans additional major meetings in the first half of next year with state governments and universities, officials say. Gibbons said that he intends to "spend more time outside the office" and to meet more scientists outside Washington in future.

Colin Macilwain

Commercial backing 'could impair academic influence'

Washington. A study of US researchers in the life sciences has warned against too high a level of industrial sponsorship for research at universities. Beyond a certain level of industrial sponsorship, academic productivity and influence appear to fall off without a commensurate increase in commercial productivity, says the study.

The study was sponsored by the US National Center for Human Genome Research, and the results were published last week in *The New England Journal of Medicine* (335, 1734–1739; 1996). Its broad message is that industry-sponsored scientists are at least as academically productive as those who do not rely on industry funding — and much more commercially productive.

But it confirms that sponsored scientists are more prone to secrecy, to choosing research topics with an eye to commercial ends, and to refusing to share materials or research results with colleagues. The authors of the study also conclude that: "Beyond some point, increases in the participation of academic institutions in relation-

of publication influence, this group was surpassed by faculty members who were not supported by industry. The highest rate of commercial outcomes, ranging from patent applications to new products and companies, came from scientists receiving between one- and two-thirds of their funding from industry.

The authors, led by David Blumenthal, chief of health policy research at Massachusetts General Hospital in Boston, write that comparison with a 1985 study shows that these broad conclusions have remained "remarkably stable" over the decade. The conclusions, they write, suggest that relationships between academic institutions and industry "enhance commercial productivity among distinguished academic investigators and do not compromise their participation in traditional academic activities".

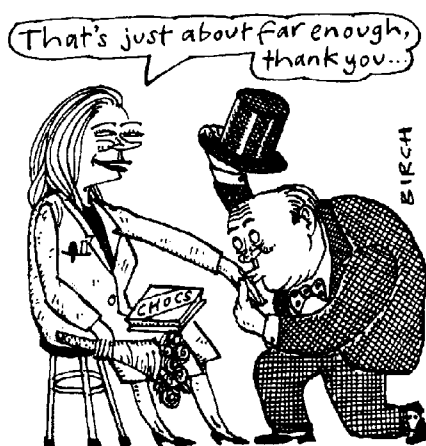
But there was a significant difference in the proportion of scientists likely to pick research topics with an eye on commercial applicability. Thirty-five per cent of industry-supported scientists said that they did this, compared with 14 per cent of academics who were not funded by industry.

Similarly, 14.5 per cent of industry-sponsored scientists said that trade secrets — defined as information kept secret to protect its proprietary value — had resulted from their work. Only 4.7 per cent of non-industry funded scientists said the same. Among industry-sponsored faculty, 11.1 per cent said that they had refused requests from colleagues to share biological materials or research results; 5.8 per cent of non-industry funded scientists said the same.

"The bottom line for me is that these relationships are working but have risks," says Blumenthal. The authors write that universities should be "vigilant" because of the greater tendency among industry-sponsored scientists to keep their results secret, and the skewing of research topics towards commercial ends.

Steven Rosenberg, chief of surgery at the National Cancer Institute, calls it "ghastly" that 14.5 per cent of industry-sponsored scientists admitted to keeping trade secrets. "We are not trying to build a better refrigerator. We are trying to save lives," says Rosenberg. "Keeping secrets is contrary to that ethic."

But industry organizations welcomed the conclusion of the study that industry-supported scientists are excelling academically and commercially. Society is "getting a good bang for its buck with the industry process," says David A Shriver, assistant vice-president for academic affairs at the Pharmaceutical Research and Manufacturers of America. **Meredith Wadman**



ships with industry may have costs that outweigh their benefits."

The study was based on information supplied by 2,052 faculty members at 50 leading US universities in 1994 and 1995. Among respondents, 28 per cent reported receiving industry support.

A limited amount of corporate funding appears linked to above-average scientific performance. Faculty members who reported receiving up to one-third of their total research budgets from industry outperformed non-industry funded colleagues in both rates and influence of their publications, as well as in their administrative involvement at their institutions.

But these effects fell off with increasing industry support, most markedly among those receiving more than two-thirds of their budgets from industry. And, in the category

US space programme 'should not centre on Mars life claims'

Washington. The US space programme should resist the temptation to overhaul completely its plans for Mars exploration to base them solely on last summer's announced finding of signs of life in a martian meteorite, a panel of the National Research Council (NRC) said last week.

The NRC's Committee on Lunar and Planetary Exploration (Complex) confirmed its support for a Mars sample return mission, but only as part of a "measured approach to the exploration of Mars" aimed at "advancing our understanding of Mars on all fronts".

"Complex believes that it is inappropriate to predicate an important aspect of future Martian studies on the unconfirmed results in a single scientific paper," wrote the panel, which had been asked by the National Aeronautics and Space Administration (NASA) to review sample return plans in the light of recent claims about life on Mars.

A programme focused only on hunting for microfossils would be inadvisable, says the report, "because unequivocal evidence may be hard to find". Instead, NASA should stay with the current strategy for Mars exploration proposed by earlier internal and external advisory groups, which begins with a global reconnaissance of the planet and includes geological and meteorological studies as well as the search for life.

The Complex panel warns that a strategy tailored only to searching for life could harm the scientific study of Mars, because "highly successful missions could be characterized as failures if they do not return with microfossils or living organisms".

The panel, chaired by Ronald Greeley of Arizona State University, prefers NASA to focus on the more comprehensive goal of "understanding Mars as a possible abode of life". Before science can settle the question of whether life exists on the planet, it must first understand how life evolved on Earth, and how planets themselves evolve. Scientists must also develop criteria for "the unambiguous identification of biotic signatures", which will require specialized equipment and laboratories, the report says.

Although the committee does not comment on specific mission scenarios for a Mars sample return, it does say that the most "aggressive" option under consideration "seems unrealistically ambitious". This requires a national commitment to Mars exploration and would land a 'robotic field geologist' on Mars as early as 2002.

This option will be scrutinized at a budget 'summit' meeting now scheduled for February at which White House and congressional leaders will try to agree on future funding levels for space.

Tony Reichhardt

Europe agrees a compromise

Paris. Consumer organizations and the food industry — but not environmentalist groups — have given a general welcome to a compromise deal in Europe on the marketing and labelling of novel foods, including genetically modified foods and ingredients. These will have to be labelled if there has been any change in their "characteristic or food property".

The agreement has been reached after five years of negotiations by a joint committee of the European Parliament and the Council of Ministers, which represents the 15 member states. The agreed text is for a 'regulation'. If approved within six weeks by both bodies, as required under European rules, its provisions will pass directly into national law.

The most significant concession won by the parliament concerns the conditions under which food is labelled. The Council of Ministers and the European Commission had wanted labelling to be required only

although the agreement falls short of what consumer organizations had wanted, it is pleased the council and commission had made significant concessions on labelling.

But not everyone is satisfied with the compromise wording. Hiltrud Breyer, a member of the parliament's Green group, describes it as a "second-best solution" and an "unsatisfactory mixture of progress and loopholes". The Green group, and other environmental organizations, argue that one loophole is that foods or ingredients identical to traditional products would not have to be labelled. It would mean, for example, that oil pressed from modified soya beans would escape the requirement for labelling, as the oil cannot be distinguished from that produced from non-modified beans.

Roth-Behrendt, who is also a lawyer, challenges the realism of such claims. She argues that labelling would be impossible to implement in practice where modified products could not be distinguished from the traditional product. "What sort of legislation would it be if it couldn't be implemented?"

In fact, she argues, the oil example demonstrates precisely the strength of the proposed regulation. She points out that, if techniques were developed that allowed oil from modified soya beans to be distinguished from that prepared from non-modified beans, it would then have to be labelled.

Another controversial aspect of the agreement is that it does not call for mixtures of genetically modified and non-modified products to be segregated and labelled. This means that mixed shipments could be imported provided they were labelled as 'possibly' containing genetically modified

FT/Colin Beere

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Chef protest: leading chefs at a Greenpeace photocall backed labelling for novel foods.

where novel foods or ingredients were "significantly different" from an equivalent existing food or ingredient.

The new wording makes the text "water-tight", claims Dagmar Roth-Behrendt (Social Democrat, Germany), the parliament's rapporteur for the committee, who argues that the term "significantly different" was ambiguous. The agreed text means that labelling will be required for all products where any difference to the traditional product can be proven scientifically, she says.

The compromise has been welcomed by both the European Consumers' Association (BEUC) and the Confederation of EU Food and Drink Industries. BEUC says that,

EU urges national

Paris. The European Commission admitted last week that genetically engineered maize has been illegally imported into the European Union (EU) since the beginning of October. One official from the commission claims that it is powerless to bring an immediate end to the imports, arguing that responsibility lies with member states.

The commission is scheduled to decide next week whether to approve Ciba's maize for import to the EU after it receives the conclusions of its three scientific committees — on foods, animal nutrition, and pesticides — as to its safety. Until the commission reaches a decision, the unprocessed maize is "automatically banned" within the EU, notes a commission official.

But in a memo sent to member governments on 27 November, the commission reveals that, according to import certificates, 4,000 to 5,000 tonnes of maize have been arriving weekly through the ports of Antwerp, Rotterdam, Lisbon and Barcelona

on food labels

material, says Steve Emmott, a spokesman for the Green group. He argues that such labels could become so widespread as to be "meaningless", so defeating the purpose of labelling.

BEUC also calls on retailers and the food industry to exert pressure on suppliers to segregate shipments. Similarly, others point out that segregation may be obtained through consumer pressure without legislation. Responding to such market pressure, major food retailers — such as the UK Iceland group — and importers have already refused all US soya bean imports. Some suppliers are beginning to offer segregated shipments.

It is still unclear how the proposed labelling requirements will relate to the rules of the World Trade Organization (WTO). Some observers suggest that they could be contested as a disguised barrier to trade. If a genetically modified product was produced by the United States but not in Europe, for example, the former could argue that labelling stigmatized its product and favoured non-modified European equivalents.

The WTO has not specifically considered genetically modified organisms, however. But, according to Behrendt, the European Parliament would resist any opposition from WTO. She argues that the legislation is non-discriminatory in that it applies to both importers and European countries. Europe has for the first time incorporated the consumer's right to be informed and to choose freely in a trade issue, she says, arguing that this "philosophy" should be respected by its trading partners.

Declan Butler

Call for reform of scientific panels

Paris. Franz Fischler, the European Union's agriculture commissioner, last week called for a radical reform of the way in which scientific advice is used in political decision-making within the union in the wake of the 'mad cow' crisis.

Testifying before the European Parliament's inquiry into bovine spongiform encephalopathy (BSE), Fischler suggested that management of the scientific committees to which the European Commission turns for advice might benefit from being transferred to independent organizations. This happens in the United States, where the government often asks the National Academy of Sciences and the Institute of Medicine to report on controversial topics.

Fischler said it is "essential" to provide the scientific advisory system with a reputation that commanded public respect, citing as a model the US Food and Drug Administration, an independent federal body but with law-enforcing powers. "As long as our scientific advice is questioned — as it has been in BSE — public confidence in our decisions is not going to be there," says one commission official.

Among the major questions that need to be addressed about the committees, said Fischler, is whether they are sufficiently independent from the interests of member states and lobby groups. The BSE issue showed that national perspectives are "not without importance", he admitted. The basis on which committee members are appointed needs to be re-evaluated, he said, as well as the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

AP/Thierry Chailier

Fischler: independent committees needed?

question of how to ensure that minority opinions are heard.

The European Parliament's inquiry is expected to produce proposals for a reform of the commission's system of scientific committees, which has come under fire following the BSE crisis. Critics have questioned both their scientific credentials and political impartiality.

Fischler pointed out that political decisions must be based on good science, because otherwise they could be contested at the European Court of Justice and the World Trade Organization (WTO). For example, under WTO rules, countries cannot ban imports or demand that they be labelled unless it can be shown that they fail to meet agreed standards, or — if such standards do not exist — that there is firm scientific evidence of a risk (see *Nature* 384, 301; 1996). This made it essential for Europe to have first-class structures for analysing and evaluating risks and their political management, he said.

D. B.

action to police imports of genetically altered maize

since 1 October. It points out that many of these shipments originate from the 1996 US harvest and are therefore likely to contain genetically modified maize, given that the United States does not segregate this from non-modified maize.

Critics point out that, because border controls between member states have been virtually abolished within the single market, maize imported by any one country can circulate freely within the rest of the EU. A spokesperson from the United Kingdom's Department of the Environment said that preventing the maize from spreading within the EU would be the responsibility of the member states whose ports receive the US shipments.

The commission seems unlikely to take any immediate action against countries importing the maize, according to one official. All it has done so far is to remind member states of their obligation to respect the law, and to carry out adequate inspec-

tions "to ensure compliance". The commission would need to first establish that the law is being broken, he explains, pointing out that the memorandum only says that "it is to be expected" that the imports contain genetically modified maize. The commission would then "get in touch with member states to see how they would rectify the situation", he adds.

"It is the responsibility of member states to comply," says the official, adding that the commission sees no point in taking action against the imports now, given that a decision on its authorization is expected later this month. "If member states are breaking regulations and the situation might be rectified two weeks later you don't drag them to court," he says.

On a wider level, the illegal imports have reopened the question of the commission's limited powers to police EU legislation. Under the union's founding treaties, it is the member states, and not the commission,

that are responsible for ensuring that EU law is implemented properly. The commission only has a role of oversight, limited to taking member states to court for having violated a treaty.

But this is a long process, while the commission's ability to prove infractions have occurred is also hampered by its limited powers of inspection. "We don't police like police in your home town do," says one commission official, "we don't arrive at the scene two minutes later." He admits that "community law is violated every day".

The issue of whether the commission should be given wider powers to police legislation has become a heated one following the 'mad cow' crisis. The commission is widely considered to have lacked sufficient powers to ensure that member states, and the United Kingdom in particular, were properly implementing EU rules on bovine spongiform encephalopathy (BSE) control measures (see *Nature* 384, 8; 1996).

D. B.

Japan faces opposition over plan to raise tuition fees ...

Tokyo. Japan's science community is expressing concern about a proposal to increase tuition fees for students studying science, engineering and medicine at Japan's national universities.

All undergraduates at the 98 national universities at present pay the same tuition fees, whatever their discipline. From 1998, however, students in science, engineering and medical departments may have to pay higher fees than other students, following a proposal by the Ministry of Finance aimed at improving Japan's deteriorating public finances.

Senior scientists and academics are concerned about the long-term impact on Japan's science and technology base. They argue that the role of national universities is to foster scholarship equally in all core disciplines. "It is a very bad policy" and shows that the ministry "does not understand the meaning of education", says Hiroyuki Yoshikawa, president of the University of Tokyo.

He and other academics argue that the proposal will impede the transfer of knowledge to the next generation and send out the wrong message to society about the importance of science and technology. Critics also argue that raising tuition fees will discourage students from taking science, engineering and medical degrees.

"We need to educate more students in

science and technology and this proposal is not healthy for Japan's future," says Akito Arima, president of the Institute of Physical and Chemical Research. The policy is also strongly opposed by the Japan Association of National Universities.

The education ministry has successfully resisted similar proposals in the past to increase tuition fees. But this time pressure from the finance ministry for the proposal to be accepted has mounted in the wake of this year's expensive bail-out by the Japanese government of failed housing loan companies, and increased costs involved in providing science, engineering and medical curricula at national universities. "The pressure is much higher" than previously, says Arima. "I hope that they can find another solution," he says.

Annual tuition fees at national universities are scheduled to rise next year by almost 5 per cent from ¥447,600 to ¥469,200 (US\$3,970 to \$4,160). Figures for the proposed extra fees that science, engineering and medical students may face from 1998 are not available.

Students at private universities in Japan, which number more than 400, already pay about 20 per cent more for science and engineering courses and more than 500 per cent more for medical and dental courses, compared with the liberal arts and humanities.

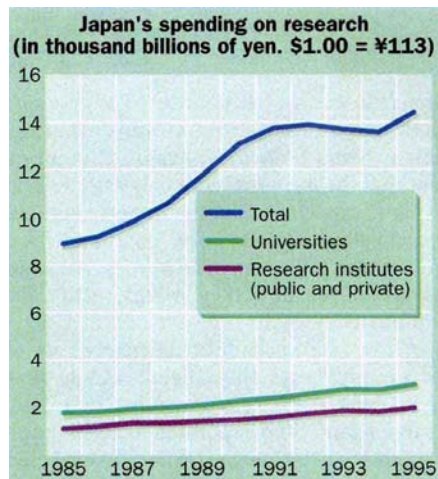
Richard Nathan

... as research funds take off again

Tokyo. Japan spent a record ¥14,408 billion (US\$129 billion) on scientific research and development in 1995, an increase of 6 per cent on the previous year, according to a survey of 15,700 organizations by the Management and Coordination Agency, which is affiliated to the prime minister's office.

The increase — the first for three years — was mainly due to improved economic conditions and two supplementary budgets from the Japanese government in 1995. The budgets contained substantial subsidies for research, designed to boost Japan's recession-hit economy, according to government officials.

Last year, Japanese universities, private and public research institutes, and companies spent the equivalent of 2.96 per cent of gross domestic product on scientific research — more than any other industrialized nation. Research spending at universities increased by 8.3 per cent and spending at research institutes by 9 per cent.



Nevertheless, almost two-thirds of scientific research spending still comes from industry. But government spending on research and development is set to rise substantially, following the approval in June this year of a new five-year plan for science and technology (see *Nature* 381, 726; 1996).

R. N.

Australian universities brace for higher fees after Senate vote

Sydney. Scientists have failed in their campaign to reverse the Australian government's decision to increase student fees for university courses. The finely balanced Senate rejected Opposition moves to oppose the increases, which were announced in the budget in August. Two independent senators backed the Coalition government, ensuring passage of the legislation.

Science-based courses are among the hardest hit, with a rise in fees of 92 per cent to A\$4,700 (US\$3,760) a year. Medicine, dentistry, veterinary science and law will suffer a rise of 125 per cent to A\$5,500. Under the Higher Education Contribution Scheme, graduates will have to pay the fees after leaving university when their taxable income reaches A\$20,700 a year. Until now the threshold has been A\$28,000.

Amanda Vanstone, the education minister, introduced the increases partly to make up for a 5 per cent cut in operating grants for universities (see *Nature* 382, 569; 1996).

University enrolments for the coming academic year have fluctuated between universities and states. Older institutions report only minor effects from the fee increases, while newer institutions are suffering drops in the number of entrants, especially for science-based courses.

The contrast is shown by comments by the vice-chancellors of the relatively new Victoria University of Technology and the older University of Sydney. A "despondent" Jarlath Ronayne at Victoria says his university has had up to 38 per cent fewer first-preference applications in science. He predicts that Australia will have to 'import' scientists — "disastrous for the future of the country".

But Gavin Brown at Sydney says science applications are 4 per cent higher. He is "relieved" that the legislation has been passed. So too is Stuart Hamilton, the new chief executive of the Australian Vice-Chancellors' Committee, most members of which had reluctantly urged the Senate to approve the measures in the interests of retaining some stability in the university system.

Vanstone describes the political victory as "a vote of confidence in Australian universities". Meanwhile, all universities are undergoing painful restructuring, with faculties, departments, courses and campuses amalgamating and shrinking as administrators come to terms with lower public funding.

Another controversial measure, approved by the Senate, allows universities from 1998 to enrol up to a quarter of their entrants from among Australian residents who can pay fees for the full cost of their courses. Universities were previously allowed to charge full fees only to students from overseas.

Peter Pockley

Radiotelescopes face costs of complexity

Green Bank, West Virginia. The US National Science Foundation (NSF) could face a \$21-million bill to cover extra costs incurred by the contractor building two long-delayed and highly complex radiotelescopes.

US astronomers are concerned that cost overruns on the telescopes will eat into NSF's budget for radioastronomy research. But NSF officials say that the delays are not unusual, considering the nature of the projects, and that it is not yet known how much extra the foundation will have to pay.

The projects concerned are a new 110-metre diameter, fully steerable radiotelescope at Green Bank, West Virginia, and a major upgrade of the largest radiotelescope in the world at Arecibo, Puerto Rico. Each is running at least two years late because of engineering problems which, officials say, reflect the complexity of the designs.

When complete, the \$75-million Green Bank instrument will be the most powerful fully steerable radiotelescope in the world (a title at present held by the Bonn telescope in Germany). Its paraboloid dish is shaped to reflect incoming radiation to a focal point beyond its edge, giving it an uninterrupted view of the sky. The \$25-million upgrade at Arecibo will increase the performance of the 300-metre dish there — fixed in the crater of an extinct volcano — by a factor of 20.

Radiation Systems Inc. (RSI) of Sterling, Virginia, now a division of Comsat, took on the projects on a fixed-price basis. But Comsat RSI is blaming the NSF observatories that manage the two contracts for contributing to the costly delays. It is suing Cornell University, which runs the Arecibo Observatory for NSF, for \$7 million for costs incurred on the upgrade.

The contractor is also claiming extra payments, believed to total \$14 million, from the National Radio Astronomy Observatory (NRAO) for cost overruns at Green Bank. NRAO is financed by the foundation. Under the terms of the Green Bank contract, the claim will go to binding arbitration if the parties cannot reach agreement. "Both projects were very much more complex than the contractor or the NSF had realised at the beginning," says Hugh Van Horn, director of the astronomy division at NSF.

Van Horn adds that the projects had taught him that "it is vital to keep a very close watch on all aspects" of such contracts "and that we cannot leave it simply to the contractor and the observatory". "We're extraordinarily pleased with how these projects are going," he says. "But there is going to have to be some manoeuvring with the lawyers before they are through." Van Horn says that the issue of finding extra money has not arisen yet and, if it does, astronomy may not bear all the cost.

The new Green Bank telescope is the legacy of Senator Robert Byrd (Democrat,

West Virginia), formerly chair of the Senate Appropriations Committee, who secured the money for it after the 1988 collapse of a previous radiotelescope at the same location. Investigators blamed metal fatigue around holes in the 20-year-old telescope's bolted structure.

NRAO decided to boost the performance of the new telescope by using an asymmetrical paraboloid dish. Designs of this type — which eliminate the shadow cast by the receiver in the centre of a standard radiotelescope disc, together with unwanted radiation scattered off the receiver legs — have been pioneered on a smaller scale by astronomers at Bell Laboratories in New Jersey.

But the shape of the dish has contributed to delays. When the project was contracted out in 1990 it was due for completion in

detect signals of wavelength as short as 3 mm. "We'll have to point it more accurately than any other telescope has been pointed."

To compensate for gravity and temperature-induced deformations of the disc, its entire shape will be controlled by actuators that can lift or lower the surface by an inch at each joint in the structure. Ground-based lasers will precisely monitor the position of each surface panel.

The NRAO is managed for NSF by Associated Universities Incorporated (AUI) in Washington. Robert Hughes, president of AUI, says that the longest delay on the contract was caused early on, when RSI had to rework its original design for the structure to meet NRAO's specification.

Hughes confirms that AUI is in discussions with Comsat RSI about its claim for extra costs, but, in common with officials at NSF and at the contractor, would not comment on its size. "There's no merit in advertising a number which could change enormously," he says. But sources in the astronomy community say the claim is for \$14 million.

Arecibo is a radar-radio telescope, one of whose uses by planetary scientists is to bounce radar beams off objects as far away as Jupiter. The upgrade will boost the performance of the telescope by a factor of 20, according to Paul Goldsmith, director of the National Astronomy and Ionosphere Center, the Cornell-managed organization that runs the observatory.

The upgrade is taking place in three phases: construction of a wire mesh fence, 15 metres high and 1,000 metres long, to protect the dish from interference, the doubling in power of the telescope's radar transmitter; and construction of a complex receiver, suspended above the dish. The 90-tonne receiver was hoisted into place in May and Comsat RSI is expected to complete work on its control systems next March — more than two years after the date that was envisaged when it was contracted to build and install the receiver in 1993. The design was already complete at that time.

Goldsmith attributes "some of the delays" to Comsat's takeover of RSI in 1994. "Project management was in turmoil," he says. McGarahan declines to comment on that, but confirms that Comsat has filed a claim of \$7 million against Cornell at the US District Court in New York.

Van Horn says that the worth of the two instruments will not be diminished by their coming into use years later than planned. "Very large instruments have a kind of decadal rhythm to them," he says. Lockman describes the delay at Green Bank as "a little frustrating. I wish we'd had it earlier, but there is nothing that bad that we've missed. I feel now that the project is getting under control".

Colin MacLiwain

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Going up: Engineering problems in upgrading the Arecibo telescope have increased costs.

1994. It is not now expected to be finished until summer 1998. Paul Vanden Baut, director of the NRAO, says that design problems, difficulties in turning the design into blueprints, and problems with assembly each contributed about a year to the delay.

A key issue, mentioned repeatedly by people close to the project, was the complexity of computer-aided design of the three-dimensional joints that hold the tubular structure of the dish together. The dish is symmetrical about one axis only: each pair of joints, facing each other across that axis, is unique. The joints are provisionally bolted together before being permanently welded.

The massive stand that will support the dish is already finished. The dish itself is two-thirds complete on the ground, with all the remaining parts delivered to the site, and dozens of welding teams working to finish the rest before winter closes in on the Appalachian mountains. Jay Lockman, chief astronomer at the site, says the big challenge for the instrument will be its performance at high frequencies, where he hopes it will

Recombinant DNA committee backs its proposed reform

Washington. The National Institutes of Health (NIH)'s Recombinant DNA Advisory Committee (RAC) on Monday gave its qualified approval to a plan for its own overhaul proposed by NIH director Harold Varmus last month. By a 12 votes to zero, with two abstentions, the committee, which currently ensures that novel gene therapy protocols are safe and ethical, endorsed Varmus's plan, which was published in the 22 November *Federal Register*.

The plan removes the RAC's role approving protocols and reduces the size of the committee from 25 to 15 members, while maintaining its job of publicly debating novel proposals. But RAC members, in a four-hour discussion, raised concerns about the weight and significance of their recommendations under the new scheme. They also expressed fears of a lack influence of proposed gene therapy policy conferences.

Varmus proposed the plan after public outcry over a proposal in July to disband the RAC altogether. He will consider the modifications suggested on Monday by the RAC in finalizing his proposal. While RAC members in general endorsed the plan, there was lengthy discussion about, and some objection to, its specifics. Some members complained that, because the FDA evaluates proposals privately, the RAC will have no way of knowing whether its recommendations on novel proposals are actually adopted. □

US global warming paper roasted

Geneva. Environmental groups reacted angrily this week to the US government's latest position paper on combating global warming, which appears to propose delaying the start of legally binding limits on greenhouse gas emissions.

In its position paper issued at the start of a two-week review meeting in Geneva of signatories to the climate convention of 1992, the United States says that short-term targets — relating to the period before 2010 — are unrealistic, and proposes that signatories be allowed to "borrow against their targets for the next period, in order to emit more in a current period". But Daniel Lashof, of the Natural Resources Defense Council, warns that this will increase pressure on developing countries to increase emissions before industrialized countries make any genuine reductions. □

European patents cheaper

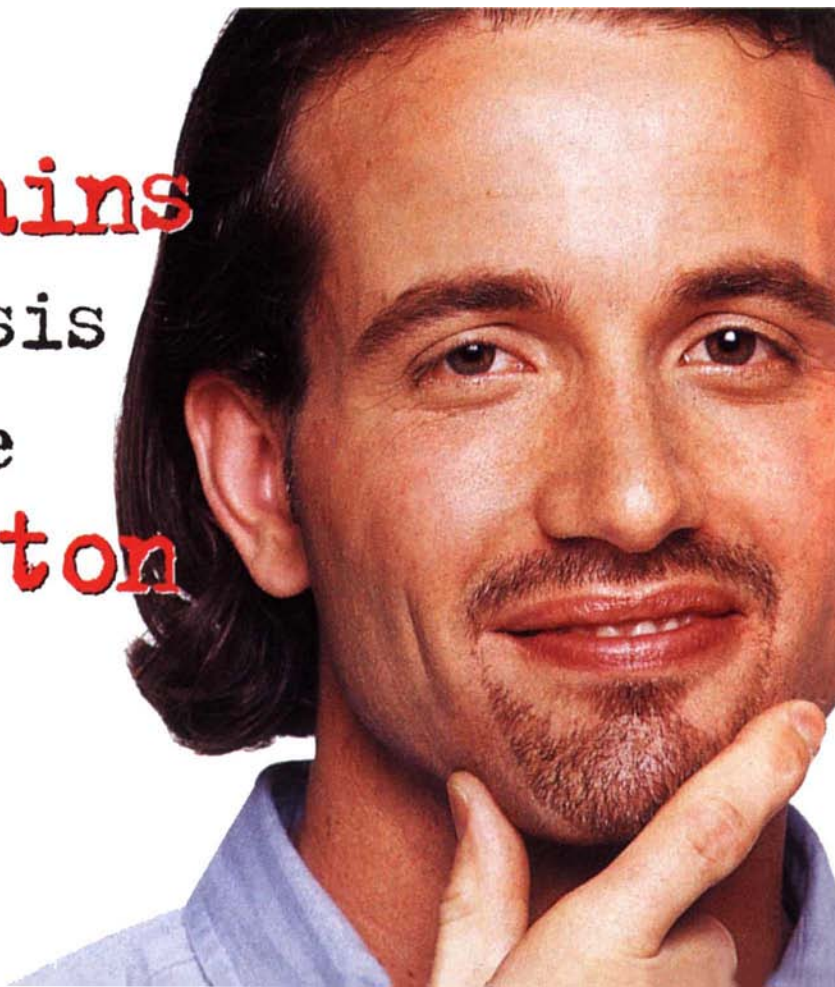
Munich. The European Patent Office (EPO) agreed last week to reduce its registration fees by 20 per cent. But as these fees are only 18 per cent of the total cost of a patent, the overall cost to an applicant — which averages DM60,000 (US\$38,960) for a patent registered in eight European countries — will fall by only 3 per cent.

The EPO is also keen to reduce translation fees, which account for more than a third of the costs of a patent, but cannot do so without changing its rules. At the moment a patent must be translated in full in all countries where it is registered, even though only the original version has legal status. Ingo Kober, president of the EPO, is proposing that the translation requirement be restricted to an abstract of the patent (unless it is being challenged). His proposal will be discussed by EPO's administrative council next year. □

Weapons plutonium disposal

Washington. The United States is planning a two-track approach to dispose of its stock of 50 tonnes of excess plutonium from nuclear weapons. Some will be burnt as mixed-oxide fuel in commercial reactors, and the rest will be vitrified for underground storage. The plan, announced last week by Hazel O'Leary, the energy secretary, marks a sharp departure from established US practice of keeping

Doug now stains
electrophoresis
gels with the
push of a button



military and civilian nuclear materials separate.

Administration officials hope that burning the plutonium in US reactors will encourage Russia to begin disposing of its own excess weapons plutonium. This position was supported recently by a US-Russian commission of leading nuclear scientists. Construction of a large plant at Savannah River, South Carolina, to convert plutonium into mixed-oxide fuel will create 10,000 jobs and help to win political support for the plan. □

Call for health research cash

Washington. The Federation of American Scientists for Experimental Biology (FASEB) has called for an increase in funding for research at the US National Institutes of Health (NIH) of 6.5 per cent in the 1998 fiscal year, which begins next October — the same increase as it successfully demanded for 1997. John Suttie, president of FASEB and professor of biochemistry at the University of Wisconsin at Madison, said that the increase, which would be worth some \$800 million to researchers, should come on top of money for a new clinical centre at NIH's campus in Bethesda, Maryland. □

Canadian vessel will resurface

Ottawa. Canadian oceanographers hope to have a replacement for the ROPOS submersible, lost in a storm in October, in the water by early next summer. This follows the approval by the Department of Fisheries and Oceans of the proposal to use insurance funds to replace the vehicle. The scientists had feared the insurance funds might be used for another purpose (see *Nature* 384, 204; 1996).

Negotiations have begun with the submersible-builder. According to Kim Juniper, secretary-treasurer of the Canadian Scientific Submersible Facility, "there is a silver lining" to the rebuilding plan because it will allow improvements to the vehicle to be made based on operational experience. □

ITER site in southern Italy?

Rome. Italy's research minister, Luigi Berlinguer, last week set up a committee to investigate possible sites in southern Italy for the International Thermonuclear Experimental Reactor. The six-member panel is expected to deliver its report before next summer.

Berlinguer wants Europe to be able to offer at least one site for the reactor when bidding to host it opens at the end of next year. Last July, Germany and France withdrew from offering possible sites. Berlinguer wants the subsidies that the European Union offers to underdeveloped regions to support a site in southern Italy. □

Lottery funds medical research

London. Britain's medical research charities are to benefit for the first time from the National Lottery. They had previously been excluded unless their work related to other approved goals, such as the relief of poverty. Of £180 million being distributed by the National Lottery Charities Board's health, disability and care programme, £6.7 million will go to medical research, of which £4.82 million is for members of the Association of Medical Research Charities. Diana Garnham, its general secretary, hoped the board would introduce an annual programme for medical research. □

French science prize awarded

Paris. Claude Cohen-Tannoudji, professor of atomic and molecular physics at the Collège de France in Paris, last week received the Médaille d'Or, France's most prestigious science prize. The prize is awarded annually by the Centre National de la Recherche Scientifique. Cohen-Tannoudji's best-known recent work is perhaps that on the use of lasers to 'freeze' atoms. This research could improve the precision of atomic clocks and allow the production of coherent atomic waves similar to conventional lasers. □

Are you spending hours at the bench staining gels? You don't have to anymore. Now you can push a few buttons, walk away and return to reproducible staining results.

Hoefer Automated Gel Stainer. the entire staining process is now automated

The new Hoefer Automated Gel Stainer makes it happen. It automates reagent delivery and development times to stain your DNA and protein gels reproducibly. It significantly reduces reagent consumption and maximizes mixing while handling your gels with a gentle rocking motion.

The Hoefer Automated Gel Stainer gives you eight pre-programmed protocols for standard silver and Coomassie blue staining—all of which can be modified to your specific needs. What's more, you can design up to 20 protocols of your own and save any of them on a removable "smart key"—keep the key to protect your protocol and simplify your start-up times in the future.

Together with our PlusOne Silver Staining Kits and Coomassie tablets, Pharmacia Biotech can provide you with everything for staining electrophoresis gels automatically. Call us: 1 (800) 526 3593 from the USA; +81 (0)3 3492 6949 from Japan; or +46 (0)18 16 50 11 from Europe and the rest of the world.

Or visit us on the Internet at <http://www.biotech.pharmacia.se>.

Ask for a free brochure; it details how you can save hours at the bench and get reproducible staining results by pushing a few buttons.



"Manual staining is an obsolete concept. It exists no more,"
Doug Burtrum,
Immunologist/Research Scientist,
Haiku writer,
New York, NY.

Pharmacia Biotech
Uppsala, Sweden. (And the rest of the world)

Divided by a common language

SIR — Koreaki Ito's suggestion (*Nature* **382**, 666; 1996) that British spellings be abandoned in favour of US spellings to assist in electronic searching of databases has led to an interesting correspondence. A solution has already been implemented, at least with some database software. Stephen Smith referred to this solution in his letter (*Nature* **383**, 384; 1996): intelligent programming.

The UNESCO CDS/ISIS software, which is widely used in libraries and museums across the world, allows us to define such a look-up table, so that whether (behaviour or behavior) or (*Nicrophorus* or *Necrophorus*) (A. T. Schafer, *Nature* **383**, 384; 1996) is keyed in does not matter. The National Information Services Corporation (NISC) ROMWRIGHT software now provides automatic searching for plural/singular forms (fish/fishes), international and other spelling variants (harbour/harbor), and different constructions for compound words (groundwater/ground water/ground-water). I am sure that there is other software that provides the same facilities. What is now needed is pressure on database hosts and other vendors to provide the same facilities in their software. This seems a much more acceptable solution than trying to impose English or American spellings (or Australian or...).

David S. Moulder

National Marine Biological Library,
Plymouth Marine Laboratory/
Marine Biological Association,
Plymouth PL1 2PB, UK
e-mail: d.moulder@pml.ac.uk

SIR — J. T. C. Sellick (*Nature* **383**, 569; 1996) has my sympathy. My family name is Ambrogio Lorenzini (without hyphen); my given name is Carlo. But *Current Contents* persists, in spite of my protests, in listing me as C. A. Lorenzini, which is both annoying and damaging. The reason is apparently that their computer software, following American usage, treats as a middle name any name coming after the given name when followed by another name.

Carlo Ambrogio Lorenzini

Dipartimento di Scienze Fisiologiche,
Universita degli Studi di Firenze,
I-50134 Firenze, Italy

SIR — I was interested in the letter from L. M. Corrochano about authors with two surnames and their treatment by bibliographic databases. The correct citation of an author's name is vital for its subsequent retrieval, and databases usually have measures in place to ensure that names of authors are cited correctly.

That is certainly the case with CAB ABSTRACTS and CAB HEALTH, the

databases produced by CAB INTERNATIONAL Information Institute. These databases cover the world literature in agriculture and some aspects of human health, and so we have rules, based on international standards, covering the citation of authors' names for many nationalities. To give a few examples, Chinese, Indian and Hungarian names present their own problems. (See also Zainab Awang Ngah, in "Malay names: current practices in ascertaining the form of headings and sources for reference", *Asian Libraries* **5** (1), 19–33; 1996.) Transliteration is necessary for names from Cyrillic and Greek scripts, and we include author variants, wherever possible, to help the searcher to retrieve all relevant papers.

For Spanish names the rule is to cite both surnames when given, even when the second one is abbreviated to an initial (Juan Carlos Palacios C., for example, is cited as Palacios C., J. C.). However, for other nationalities where we are sure that they are surnames the rule is to cite only the last of the names, thus Sellick, J. T. C. Where they are hyphenated, both names are cited, thus Clark-Sellick, J. T.

Corrochano suggests that the option of hyphenating surnames keeps journal editors and database curators happy. It should also keep the authors themselves and database users happy, if the names are cited properly.

Mike Hails (Hails, M. R.)

CAB INTERNATIONAL,
Wallingford, Oxon OX10 8DE, UK

Pump priming works in Singapore

SIR — Terence Kealey (*Nature* **383**, 474; 1996) claims that because of the establishment in Singapore of the National Science and Technology Board (NSTB) and the National Technology Plan in 1991, the rate of growth in the total research and development (R&D) budget fell after 1992. He claims that this is consistent with the laws of scientific research that he published in his book *The Economic Laws of Scientific Research*.

The effectiveness of NSTB can be measured most accurately by comparing the period just before and after its formation.

NSTB was formed in 1991 to promote economically relevant R&D in Singapore with the aim of enhancing our economic competitiveness. Before the formation of NSTB, from 1987 to 1990, GERD (gross expenditure on R&D) was growing at a compound annual growth rate (CAGR) of 15% (see table). However, since the forma-

tion of NSTB in 1991, GERD growth from 1990 to 1995 had a CAGR of 19%. This is more than double the GERD registered five years ago. Therefore, the rate of GERD growth since the formation of NSTB has been more significant in the past five years, compared to the few years before its formation.

Second, private sector GERD grew at a CAGR of 11% from 1987 to 1990. However,

	GROWTH OF ECONOMICALLY RELEVANT R&D	
	1987–90	1990–95
Gross expenditure on R&D (GERD)	15%	19%
Private sector expenditure on R&D	11.1%	23.3%

er, between the formation of NSTB in 1991 and 1995, private sector GERD has been growing at a CAGR of 23% annually. So the rate of growth of private sector GERD after the formation of NSTB is more than double that in 1987–90. Furthermore, private sector GERD growth since 1990 (at 23% CAGR) even exceeds that for overall GERD, which grew at a CAGR of 19% annually. So government funding had not only not displaced private sector expenditure on R&D, but it had also stimulated greater R&D growth in the private sector. In fact the share of private sector GERD increased from 54% in 1990 to 64.5% in 1995.

Lee Kheng Cheok

Planning and Policy Division,
National Science and Technology Board,
10 Science Park Road, The Alpha,
Singapore Science Park II,
Singapore 117 684

No relation

SIR — Philip Anderson's otherwise excellent obituary of Sir Nevil Mott (*Nature* **383**, 121; 1996) quite erroneously makes me Mott's nephew by marriage. There is no such relationship.

Charles Frank

Orchard Cottage, Grove Road,
Coombe Dingle, Bristol BS9 2RL, UK

Nothing new

SIR — Doo Jung Jin (*Nature* **383**, 662; 1996) would rename the kilogram "einstein" on the grounds that it is the only SI basic unit that is a multiple of another unit. Is (s)he aware that the einstein is a photochemical unit corresponding to Avogadro's number of photons?

Gerald E. Smyth

Nippon Shinyaku Co. Ltd,
Nishiohji-Hachijo,
Minami-ku,
Kyoto 601, Japan

Paranoid about peer review?

SIR — During my six years as a postdoc in England, I published a reasonable number of papers without ever having any reason to question the fairness of the peer-review system used by scientific journals. The papers just went through, some came back for alterations and that was that. Upon my return to Brazil, however, I began to experience more difficulty in getting my papers accepted by the same journals. Nowadays a manuscript almost always goes round at least twice before it gets published.

Except perhaps for the speed with which the data are generated, neither the style nor quality of the papers seems to have changed. My group continues to produce original work, scrupulously conducted and genuinely contributing to the general discussion on the theme, which has not changed either.

Could it be that papers originating from research institutions in 'developing countries', to use an accepted euphemism, are more rigorously reviewed than those from laboratories elsewhere? I think they are. This almost inescapable conclusion is reinforced by two major observations based on a pattern detected in the letters from the referees. First, they criticize the quality of the results, as in clarity of protein/DNA gels and so on, although I find no differences between what we produce and what is presented in most journals. These criticisms are usually final and leave no room for an appeal. Second, whenever there is a conflict between the referees, the editor's final decision invariably tilts towards a recommendation for rejection, rather than getting a third opinion. This happens even when the unfriendly referee does not present weighty arguments, or is obviously not familiar with the subject.

The question now arises as to the existence or not of such bias. How could one, in a true scientific manner, test this hypothesis? Separating the objective from the subjective is very difficult, but as the scientific publishing business is entirely based on peer review, it is surely worthy of discussion. I have racked my brain to come up with an experimental protocol that could confirm or dismiss my suspicions. The answer is that, short of resorting to fabricating different addresses for the same manuscript, an obviously objectionable practice, there is no simple answer.

On the other hand, if the much more esoteric problem of establishing the veracity of the Indian rope-trick can be tackled (*Nature* 338, 212–213; 1996), analysing prejudice in reviewing papers should be comparatively easy to establish. Taking advantage, therefore, of the fact that *Nature* is a high impact journal with a worldwide readership, it should be feasible and interesting to ask whether my Southern Hemisphere colleagues share my predicament. Surely the resulting opinion poll would be amenable to

a proper statistical treatment? Only then would it be revealed whether the tropics have turned me into a bad scientist or just a paranoid.

Franklin D. Rumjanek

Departamento de Bioquímica Médica
ICB/CCS,

Universidade Federal do Rio de Janeiro,
Ilha do Fundão, CEP 21941-590,
Rio de Janeiro, Brazil

South African priorities

SIR — As a former policy adviser to the Department of Arts, Culture, Science and Technology, I should like to take issue with a number of the points raised in your recent Briefing on science in South Africa (*Nature* 384, 11–15; 1996). In particular I feel that the article fails adequately to reflect the inherently weak position from which the scientific community starts in its attempts to achieve extra funding. This has little to do with the political allegiance of the minister concerned, but more with perceptions of what South African science has delivered in the past, and what it might deliver in the future. To imagine that the Ministry of Finance would reverse its policy on tax breaks at the behest of the science community is wishful thinking indeed.

Comments about South African science based on citation counts miss the important fact that some citations are inevitably inflated by publications on research aspects of South Africa's natural environment. It is obviously easier for a local researcher to write a paper on a unique indigenous species (and of course to research it well) than someone abroad.

Furthermore, while it is true that science funding of \$23 per capita is high for a newly industrializing country, when corrected for the fact that this science is conducted by the 'first world' component of the country, spending is actually commensurate with that of Australia, Spain or New Zealand. When the scientific community is asked how it performs in comparison to those countries, the answer is: not very well.

Ultimately the funding debate will turn — perhaps erroneously — on perceptions of value for money. The department does not need convincing about the need to identify priorities, as it has advocated this since its establishment. In contrast, those who need to absorb this message are the universities, which cling to notions of academic freedom, and research councils, which are trying to prove that their activities are already socially attuned.

Michael Kahn

PO Box 1833,
Houghton 2041, South Africa

Mote and beam

SIR — David Edge, the Editor of *Social Studies of Science* complains (*Nature* 384, 106; 1996) of recent attacks against the sociology of science using rude language and intemperate tone. He should have said more precisely against the 'sociology of scientific knowledge' (SSK) the fashionable school of sociology of science that came to dominate his journal. There were no such attacks against the older Mertonian school.

Browsing through the 25-year-old volumes of *Social Studies of Science* (in which the following references appear), I have found plenty of anti-scientist polemics and intemperate language in papers by SSK-ists. For example S. Fuller called scientists "knowledge-mongers" (19, 397; 1989) who deal with a "corrupt realm of privileged objects". Steven Weinberg, the Nobel-prizewinning theoretical physicist and science administrator was called a "megalomaniac" (24, 157; 1994). Another Nobel prizewinner, Herbert Simon, was characterized as revealing "a muddled grasp of the metatheoretic issues surrounding his own work" and accused of joining the dispute on computer simulation of discovery because "it must be grant renewal time" (21, 149; 1991). A similar charge was addressed to Weinberg and his "elite scientist friends" for using rhetoric to justify spending billions of dollars. Noam Chomsky, the world-famous linguist, was called an "upstart linguist" and blamed for "cognitive chauvinism" (19, 629; 1989).

Scientists are like "high priests" (14, 483; 1984), they present themselves like God but they are more like the devil (the latter is not original: William Blake was the first to depict Newton as the devil!). Another author characterizes the opposing camps as "us", social scientists = civilized human beings, versus "them", knowledge engineers = primitive savages (24, 106; 1994). Edge talks about close collaboration with scientists. He has a point: unlike Merton, Kuhn and Popper, the SSK-ists did go into the laboratories, did observe how everyday science was carried out and to that extent collaborated with the scientists. As elsewhere in anthropology, it would be too much to expect contributions by the 'natives' in the opinions expressed.

There was one occasion when a scientist was given a voice explicitly; Jonas Salk wrote the introduction to the book by B. Latour and S. Wolgar (*Laboratory Life: The Social Construction of Scientific Facts*, Princeton, 1986). He expressed his doubts about SSK and stressed that many details of his own work did not fit. Such opinion would probably be typical of those scientists who collaborated in the SSK field-work. Perhaps it is time to stop the unscholarly behaviour on both sides.

George Magyar

31 South Avenue,
Abingdon, Oxon, UK

The Great Pyramids of Canada

SIR — High unemployment is often conceded to be a pressing social problem, but its costs can be hard to visualize. The unemployed can become practically invisible to the wider society, and the social costs in crime rates, health expenditure and marriage breakdown are masked by the influence of other variables. But Ian Stewart's comments¹ on Stuart Wier's study² of the workforce needed to build the Great Pyramid of Khufu at Giza suggests a way of visualizing the magnitude of unemployment.

Suppose that, instead of sitting around feeling useless, the unemployed were *doing* something useless — such as building pyramids using only hand labour. Minimal capital equipment (for example chisels) would be required and the subsistence costs of the unemployed would be borne. Measured in 'pyramid equivalents', or 'khufus', how large is excess unemployment in the 1990s?

There is always *some* unemployment — so one should not count all of the unemployed as potentially available for pyramid construction. If we take Canada as an example, a conservative estimate of the 'natural' rate of unemployment would be 7.5% (the actual rate in 1989). Unemployment in the 1990s has been substantially higher (11.3% in 1992, 10% in October 1996), largely because monetary policy has aggressively pursued zero inflation, using high interest rates to restrain aggregate demand. The cumulative difference between actual unemployment from 1990 to 1996 and a 7.5% rate of unemployment amounts to approximately 2.7 million person-years of potential work — on average, about 450,000 person-years annually.

How many khufus is this equivalent to? Wier estimates that between 8,300 and 10,600 men working full-time for 23 years could have built the Great Pyramid. If one takes the higher estimate (243,800 person-years), and assumes a five-day working week, a month's vacation and ten holidays per year (implying 223 working days, rather than 365), a total of 379,500 person-years would be needed. Even without modern machinery, fewer people would probably be required today, as the Canadian unemployed of the 1990s are significantly better nourished, and of larger stature, than labourers of 4,600 years ago.

Furthermore, as Wier points out, "the manpower for moving stone blocks from the quarry to the point of installation is the largest single factor in pyramid construction", and presumably in the 1990s one might not choose to use the wooden sleds lubricated by clay and water that ancient Egyptians had to pull.

If, however, historical authenticity were the determining factor in choice of construction method, Canada's excess unem-

ployed could have built about 1.2 khufus annually, or seven between 1990 and 1996. Lined up in a row, this would surely be an impressive memorial to the economic policies of the 1990s.

Lars Osberg
Dalhousie University,
6220 University Avenue,
Halifax, Nova Scotia B3H 3J5,
Canada
e-mail: osberg@is.dal.ca

1. Stewart, I. *Nature* **383**, 218 (1996).

2. Wier, S. *Camb. Archaeol. J.* **6**, 150–163 (1996).

Theoretical papers

SIR — I should like, from my own experience working in ecology, to draw attention to problems with the refereeing of theoretical papers. Comments by Fred Hoyle¹ in his biography suggest that similar problems exist in other areas of science.

Three of my papers were rejected by the first journal to which they were sent, but then accepted by a second journal of comparable quality. The first paper² was a reanalysis of previously published data using multivariate methods, and the other two were straight ideas papers^{3,4}. These latter two were rejected by the first journals (in one case apparently without review) but received encouraging referees' reports after submission to the second journals, in one case with an invitation to produce a longer paper dealing with related topics.

As has been pointed out by the editors of a leading ecology journal⁵, the problem with many innovative theoretical papers is to distinguish between papers that are ahead of the field (tomorrow's classic papers) and those that are simply wrong or unoriginal.

Two morals can be drawn from this story. First, theoretical papers should receive lighter refereeing than data-based papers, because it is more important that an interesting idea gets into print than it is to exclude poor ideas. Errors in a data paper may be misleading to future workers, whereas most scientists rightly treat the work of theoreticians with some scepticism and are therefore less likely to be misled by a poor theoretical paper. Theoretical papers probably receive harsher refereeing at present than do data-based papers.

Second, young theoretical scientists should not take a rejection as final. If they believe their work is good, they should resubmit it to another journal. The second

set of referees' reports are often so different from the first that they might be thought to refer to a completely different paper.

David M. Wilkinson
Biological and Earth Sciences,
Liverpool John Moores University,
Byrom Street, Liverpool L3 3AF, UK

Fact and fission

SIR — C. F. v. Weizsäcker¹ quotes Lise Meitner as saying that the idea of "radium isotopes" being obtained from slow neutron bombardment of uranium was "nonsense". According to Weizsäcker, several radium isotopes were reported by Hahn and Strassmann² when they checked a finding by Joliot (although this appears to be a myth, as Hahn and Strassmann make no mention of the Joliot work in their report). The appearance of radium isotopes led to a paradox: it appeared to require emission of α -particles from uranium bombarded by slow neutrons with energies well exceeding those of the naturally emitted α -particles, in contradiction of earlier experiments in which no evidence for such "long-range" α -particles was found^{3,4}.

Scientists who disbelieve a finding and have the means at their disposal to disprove it usually try to do so. As Meitner stayed in touch with Hahn after she fled from Berlin⁵, her scepticism may have spurred Hahn and Strassmann to attempt to disprove the evidence for the radium isotopes, and in this they succeeded. In a contrasting situation, when Irène Curie and Savitch⁶ found an indication of an isotope of lanthanum having a half-life of 3.5 hours in the "pre-fission" days, it was Hahn who disbelieved the evidence and is reported to have teased Joliot that he should teach his wife chemistry⁷. Hahn and Strassmann then tried to disprove the existence of the lanthanum isotope, but instead confirmed it.

Important discoveries are often surrounded by myths. We now learn that Hahn telephoned Weizsäcker to tell him of the discovery of barium. This destroys at least the myth³ that Hahn did not tell the Berlin physicists of his discovery before publication while informing Meitner, thus giving her and Otto Frisch an "unfair" advantage in proposing the fission model!

M. Goldhaber
Department of Physics,
Brookhaven National Laboratory,
Upton, New York 11973, USA

1. v. Weizsäcker, C. F. *Nature* **383**, 294 (1996).

2. Hahn, O. & Strassmann, F. *Naturwiss.* **26**, 755 (1938).

3. Chadwick, J. & Goldhaber, M. *Proc. Camb. Philos. Soc.* **31**, 612 (1935). T

4. Hahn, O., Meitner, L. & Strassmann, F. *Z. Physik* **106**, 249; 1937).

5. Sime, R. L. *Lise Meitner: A Life in Physics* (University of California Press, 1996).

6. Curie, I. & Savitch, P. *J. Physique et Radium* **9**, 355 (1938).

7. Recalled by Bertrand Goldschmidt at the 50th anniversary commemoration of the discovery of fission in Washington DC in 1989.

1. Hoyle, F. *Home is Where the Wind Blows* (University Science Books, California, 1994).

2. Wilkinson, D.M. *Europ. J. Protistol.* **26**, 117–121 (1990).

3. Wilkinson, D.M. & Dickinson, N.M. *Oikos* **72**, 298–300 (1995).

4. Wilkinson, D.M. *J. Biogeogr.* **23**, in the press (1996).

5. Malmer, N. & Enckell, P.H. *Oikos* **71**, 171–176 (1994).

When Galileo met Ganymede

David J. Stevenson

Jupiter's satellite Ganymede has a magnetic field. Its cause could be an internal dynamo that was kick-started a billion years ago when the satellite was temporarily pushed into a more eccentric orbit.

IN 1610, in one of the most dramatic early observations stemming from the invention of the telescope, Galileo Galilei discovered Io, Europa, Ganymede and Callisto — these, in ascending order of distance from the planet, are the four largest satellites of Jupiter. At the moment, NASA's Galileo spacecraft is on an odyssey that brings it close to each of these Galilean moons. The spacecraft's findings may not be as revolutionary as those of Galileo himself, but they are nonetheless full of surprises and stimulation for planetary scientists. By this measure, the Galileo mission¹ has already been a huge success.

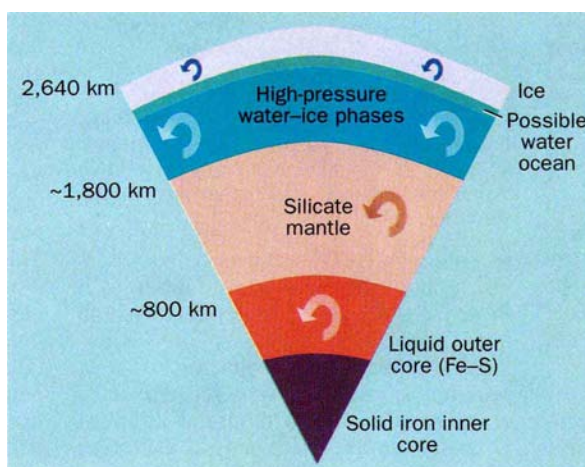
Four papers in this issue²⁻⁵, beginning on page 535, describe some of the mission's latest discoveries. They concern Ganymede, which with a radius of about 2,640 km is Jupiter's largest satellite. Most remarkably, it seems that Ganymede has a substantial magnetic field, putting it in select company in the Solar System.

To appreciate the significance of this result it is necessary to understand some things about icy bodies and about planetary magnetic fields. Ganymede is the largest member of a class of solid bodies in the Solar System that are characterized by a large amount of water ice. This class is dominated by satellites of the giant planets, such as Titan (Saturn) and Triton (Neptune), but also includes the planet Pluto. These bodies formed beyond the asteroid belt, which lies between Mars and Jupiter. Outside the belt the most common solid material is water ice but the rocky constituents found in terrestrial bodies are also present. Ganymede is thus an aggregation of materials similar to those that make up Earth, augmented by about as much water ice by mass (three times as much by volume).

When images of Ganymede were returned by the Voyager spacecraft about 17 years ago, it was evident that Ganymede's icy surface is not merely affected by impacts but also has 'grooved' terrains which appear to have been formed by internal processes. In consequence, it has long been thought that Ganymede is differentiated — that is, a body in which the materials heavier than

water ice have settled to form a core.

Galileo has now confirmed this view. The gravity results of Anderson *et al.* (page 541)⁴, which were obtained by analysing the acceleration of the spacecraft near its closest approach to Ganymede, show that the satellite has only about 78 per cent of the moment of



Inside view — a cross-section of Ganymede, showing its various layers as inferred in part from the new data reported in this issue²⁻⁵. All regions are convecting except perhaps the very outermost part of the ice shell and the innermost solid iron core. Ganymede's magnetic field may arise by dynamo action in the liquid outer core and be about one gauss there, similar to the field at the Earth's surface.

inertia of a body of uniform density, which makes it the most centrally concentrated solid body thus far encountered in the Solar System. Indeed, it seems that Ganymede is rather like Io, which has a radius of about 1,800 km, but with an 800-km layer of ice on top.

Earlier gravity data for Io⁶, also obtained by Galileo, show that this rocky body has an iron core. Is the same true of Ganymede? This is where the magnetic-field observation comes in. As the results of Kivelson *et al.* show (page 537)³, the magnetometer instrument detected clear evidence for a large increase in field strength at Galileo's closest approach to Ganymede. These data, and the conclusion that Ganymede has a substantial magnetic field, find strong support in the observations of Gurnett *et al.* (page 535)², which show plasma-wave radio emissions strikingly similar to those observed in the magnetospheres of magnetized planets.

At Ganymede's surface, the magnetic

field is only about a tenth as strong as Earth's. But that makes it much larger than the field at the surface of Mercury, Mars, Venus or Earth's Moon. The question of its cause is discussed by Schubert *et al.* (page 544)⁵, who conclude that it is unlikely that the field is due to permanent magnetism. Rather, it would seem that Ganymede has a dynamo, just like Earth.

Dynamos operate by induction: conducting fluid moving relative to the field generates an electromotive force which drives the currents responsible for the field. At first sight, the idea of a dynamo inside Ganymede's iron core seems reasonable. After all, there is good evidence that Io has an iron core; and Io possibly has a dynamo (ref. 7; see also Schubert *et al.*⁵), although it is difficult to tell because Jupiter's field and plasma effects are so large. Moreover, it is likely that such a core would be in a liquid state, because of the low freezing point of an iron alloy containing large amounts of sulphur, and there is ample evidence for sulphur in Jupiter's system. The low melting point also makes it easier to form a core.

The puzzle comes when one thinks about how sufficiently vigorous motions within that core might arise. Convection of some kind seems to be the only plausible cause, and could stem from the steady cooling of the core over geological time. Our current understanding of terrestrial bodies⁸ (those within the asteroid belt, including Earth) suggests that a solid-iron inner core helps in generating convection because growth of this core during cooling allows sulphur and other light elements to rise up, releasing gravitational energy and providing buoyancy for fluid motions. However, bodies as small as Ganymede are unlikely to have sustained the necessary fluid motions for billions of years, because the heat can escape by thermal conduction, so the puzzle remains.

Perhaps the answer lies in tidal heating — a process that arises in these satellites when the orbit is substantially eccentric and the solid material within the satellite is periodically flexed. (This is just like the heating you observe when you bend a

piece of metal or a plastic credit card back and forth.) Ganymede, unlike Io and Europa, does not currently undergo much tidal heating, but may have done so in the past. Malhotra⁹ and Tittmore¹⁰ have shown that Ganymede may have passed through an orbital resonance as the orbits of the three innermost Galilean satellites expanded under the action of tides raised by Jupiter. In the more plausible of these schemes, the one proposed by Malhotra⁹, Ganymede may have passed through a relatively recent resonance (perhaps a billion years ago) in which its orbital eccentricity was pumped to a high value. The same tidal flexing that drives volcanoes on Io and resurfacing on Europa may then have revitalized Ganymede's interior^{11,12}.

Through the promotion of additional differentiation triggered by tidal heating, or merely by heating of the core, Ganymede's thermal-evolution clock may have been reset so that it now behaves like a body that is merely a billion years old rather than its true age of 4.5 billion years. Interestingly, this view may be consistent with the evidence of low crater densities on some terrains on the surface of Ganymede. Gene Shoemaker (Lowell Observatory) has suggested that the youngest terrains on Ganymede may be much younger than we supposed a decade ago — perhaps younger than a billion years.

Whatever the explanation for Ganymede's field, its discovery is an exciting addition to knowledge about cosmic bodies with magnetic fields. It may also provide clues to the evolution of these bodies. Moreover, it endows Ganymede with a phenomenologically rich environment, because the satellite's magnetosphere influences the particle bombardment of the surface and the behaviour of its tenuous atmosphere. Comparison with Callisto (data currently under analysis) and Europa (first fly-by scheduled for 19 December) will add to the emerging picture of this 'planetary' system first found almost 400 years ago. □

David J. Stevenson is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

1. <http://www.jpl.nasa.gov/galileo/>
2. Gurnett, D. A., Kurth, W. S., Roux, A., Bolton, S. J. & Kennel, C. F. *Nature* **384**, 535–537 (1996).
3. Kivelson, M. G. et al. *Nature* **384**, 537–541 (1996).
4. Anderson, J. D., Lau, E. L., Sjogren, W. L., Schubert, G. & Moore, W. B. *Nature* **384**, 541–543 (1996).
5. Schubert, G., Zhang, K., Kivelson, M. G. & Anderson, J. D. *Nature* **384**, 544–545 (1996).
6. Anderson, J. D., Sjogren, W. L. & Schubert, G. *Science* **272**, 709–712 (1996).
7. Kivelson, M. G. et al. *Science* **273**, 337–340 (1996).
8. Stevenson, D. J. *Rep. Prog. Phys.* **46**, 555–620 (1983).
9. Malhotra, R. *Icarus* **94**, 399–412 (1991).
10. Tittmore, W. C. *Science* **250**, 263–267 (1990).
11. Showman, A. P. & Malhotra, R. *Icarus* (submitted).
12. Showman, A. P., Stevenson, D. J. & Malhotra, R. *Icarus* (submitted).

From the nose to the brain

Cori Bargmann

Do two people perceive the same thing when they say that jasmine smells floral? In other words, is information from the olfactory system organized similarly in different individuals? Mombaerts and colleagues argue that it is, as described in a paper in *Cell*¹ last month. They have used elegant mouse genetic methods to tackle the question, and the results raise the intriguing possibility that the olfactory

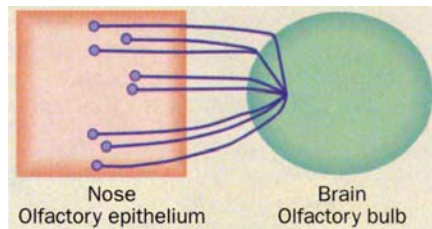


FIG. 1 Olfactory neurons expressing a particular receptor gene are scattered through a region of the olfactory epithelium. Their axons converge on a few target glomeruli in the olfactory bulb.

receptor proteins themselves help to organize a map of olfactory information as it flows from the nose to the olfactory bulb in the brain.

In the mammalian olfactory system, odorants are thought to be recognized by a family of around a thousand different G-protein-coupled receptors². Different odorants activate different subsets of receptors, allowing discrimination between an enormous variety of odorants. Each receptor gene is expressed in a small fraction of olfactory neurons, and it is likely that each neuron expresses only one or a few receptor genes^{3,4}. The olfactory neurons expressing a given receptor are scattered across a large region of the sensory epithelium, and their axons each project to the olfactory bulb (Fig. 1).

Imaging populations of neurons expressing particular receptor messenger RNAs showed that axons expressing one receptor gene converge on a few small target regions in the bulb^{5,6}. Interestingly, the target regions (glomeruli) for any one receptor are at similar positions in different mice, suggesting that the pattern of axon convergence is fixed among all individuals. A developing olfactory neuron makes two precise choices from a vast repertoire: first it selects one receptor from a thousand possibilities, and then it finds one target region out of a thousand.

To extend these results, Mombaerts et al.¹ developed high-resolution genetic methods to image the axons of individual olfactory neurons expressing a specific receptor gene. A gene fusion between lacZ and the microtubule-associated protein tau⁷ was coexpressed with the P2 olfactory

receptor gene to permit blue β -galactosidase staining of the P2 subset of olfactory axons. It is completely unclear how an olfactory neuron chooses to express a single receptor gene such as P2, so the tau-lacZ gene was inserted within the genomic copy of the P2 gene by a homologous recombination 'knock-in' strategy. The reporter gene was engineered so that a bicistronic mRNA would express both the reporter and the endogenous P2 gene, causing minimal disruption to the neurons (Fig. 2). This method should be generally useful for examining projections of neurons that express a given gene.

As expected from the previous studies, the blue axons expressing P2 and the tau-lacZ reporter converged onto a few glomeruli — one glomerulus each on the medial and lateral halves of each olfactory bulb — and the single-axon resolution of the lacZ staining revealed that all of the axons expressing P2 converged on these glomeruli. The positions of the blue glomeruli were constant in different mice. Thus, a few defined glomeruli within each mouse receive all inputs from a single olfactory receptor.

Sensory systems use both developmentally hard-wired strategies and activity-dependent mechanisms based on correlated firing to segregate axons at their targets⁸. Although correlated activity could con-

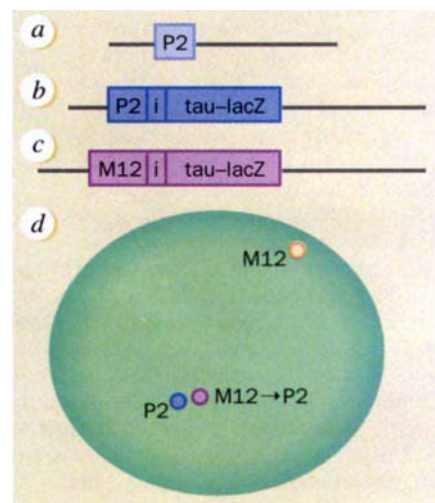


FIG. 2 Gene targeting at the P2 olfactory receptor locus. a, The normal locus. b, The targeted locus expressing both P2 and a tau-lacZ fusion protein at the same locus due to the insertion of an internal ribosome entry site that creates a bicistronic locus. c, The precise replacement of P2 coding sequences with those of the receptor M12 within the P2 genomic locus. d, The site of convergence within the olfactory bulb of blue (lacZ-positive) axons for the normal P2 locus, the normal M12 locus and the M12 + P2 hybrid gene shown in c.

tribute to convergence of the olfactory axons expressing the same receptor, it is unlikely that correlated firing carries enough information to create a precise map matching a thousand different receptor genes to a thousand different glomeruli in a reproducible fashion. So the basic olfactory map is probably established by a developmentally hard-wired strategy.

What molecules direct the specific guidance of olfactory axons to their target glomeruli? The only known difference between different olfactory neurons is the expression of different receptor proteins. To find out whether the receptors themselves contribute to target selection⁹, Mombaerts *et al.* genetically altered the expression of two receptors. Axons expressing the olfactory receptor M12 converge on glomeruli far from the P2 glomeruli. The coding region of the endogenous P2 gene was replaced by the coding region of M12 by recombination, causing these cells to express a hybrid gene with the M12 receptor and tau-lacZ under control of the P2 promoter (Fig. 2). The blue axons converged on glomeruli that were distinct from either the normal P2 glomeruli or the normal M12 glomeruli, but very close to the normal P2 glomeruli.

So the P2 and M12 receptors do influence the targeting of axons in the olfactory bulb, but their effect is subtle. Could the effect of receptor replacement be solely due to differences in neuronal activity conferred by expression of M12 instead of P2 receptors? The M12→P2 hybrid gene converges on glomeruli that are in a stereotyped position relative to the authentic P2 glomeruli, so in this model activity-dependent processes would have to result in reproducible patterns of segregation over small distances. Such instances have been described, but are relatively rare¹⁰.

The tantalizing alternative explanation is that olfactory receptor proteins on developing axons interact with molecules in the olfactory bulb to help guide the axons. Because the effects are subtle, there may only be a few distinct classes of receptors: it is not necessary to invoke a thousand receptors with a thousand different guidance codes. The bulb is unlikely to produce specific odorant molecules, but the olfactory receptors might recognize other

proteins, or homophilic interactions among the receptors might influence the relative arrangement of the axons as they converge. To test this idea, it will be important to determine whether the receptor is actually present on axons, and whether the effect results from differences between the P2 and M12 receptor proteins, differences between the expression levels of P2 and M12, or differences between P2 and M12 DNA or RNA sequences.

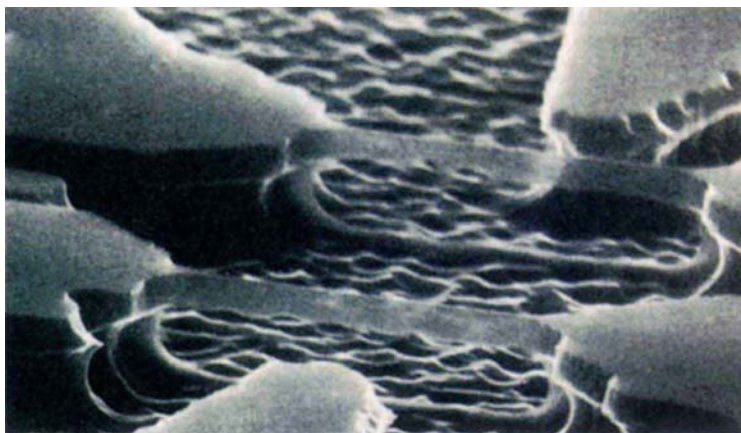
A neuron expressing the M12 receptor in the P2 chromosomal locus is much more like a P2 neuron than an M12 neu-

ron in its targeting. Therefore, guidance molecules other than olfactory receptors are the main cue distinguishing axons of P2 and M12 neurons. The reproducible olfactory map in the bulb must arise from coordinated expression of particular olfactory receptors with these unknown guidance molecules; this constant map could provide a substrate for shared olfactory perceptions among individuals. □

Cori Bargmann is in the Howard Hughes Medical Institute, Department of Anatomy, University of California at San Francisco, San Francisco, California 94143, USA.

MICROENGINEERING

Narrow in the beam



BRIDGES for bacteria? Not quite. Tuning forks for bacteria would be a little closer. These undercut silicon beams are so small (about eight micrometres long) that they vibrate more than 70 million times a second — a note about 17 octaves above middle C. The fabrication method used to make them, described by Cleland and Roukes in *Applied Physics Letters* (69, 2653–2655; 1996), should be capable of producing even smaller beams, with record frequencies of a few gigahertz (10^9 Hz). At low enough temperatures, such tiny oscillators could behave in a detectably quantum-mechanical way.

A single crystal of silicon was sculpted in several stages, involving photolithography, oxidation, metal deposition and etching, with the layers of silicon oxide and of nickel acting as masks for the etching stages. It is a relatively simple processing sequence, and has another important advantage over old methods: the final, vital undercutting stage is performed by a dry etch instead of the usual wet-etch process that can damage structures by surface tension.

To pluck the beam, or perhaps rather to bow it, the authors deposit a layer of gold on the top surface of the silicon

and drive an alternating electrical current through it. With a magnetic field imposed perpendicular to the beam by a solenoid, the alternating current forces the beam to vibrate up and down — strongly so at the resonant frequency — and that movement is detected as an induced voltage.

Free-standing structures on this scale are remarkable enough. These beams are so small that they probably contain no lattice defects, so experiments can be made on the true properties of bulk silicon; and similar devices might be used as an entirely new type of particle and energy sensor. But because the period of vibration depends on length squared, a beam just 2 μm long would be a gigahertz oscillator — and then things could get even more interesting.

Such an oscillator might be able to detect a single thermal phonon (a quantum of internal heat). Cooled down enough, to reduce interference from its environment, one could even be put into a quantum superposition of vibrational states, opening up a whole new area of experimentation. It also raises an intriguing, if remote, possibility: the science of quantum computing is still very young, but might we see a quantum-mechanical difference engine one day?

Stephen Battersby

1. Mombaerts, P. *et al. Cell* **87**, 675–686 (1996).
2. Buck, L. & Axel, R. *Cell* **65**, 175–187 (1991).
3. Ressler, K. J., Sullivan, S. L. & Buck, L. B. *Cell* **73**, 597–609 (1993).
4. Vassar, R., Ngai, J. & Axel, R. *Cell* **74**, 309–318 (1993).
5. Ressler, K. J., Sullivan, S. L. & Buck, L. B. *Cell* **79**, 1245–1256 (1994).
6. Vassar, R. *et al. Cell* **79**, 981–992 (1994).
7. Callahan, C. A. & Thomas, J. B. *Proc. Natl Acad. Sci. USA* **91**, 5972–5976 (1994).
8. Goodman, C. S. & Shatz, C. J. *Cell* **72**, Suppl. 77–98 (1993).
9. Singer, M. S., Shepherd, G. M. & Greer, C. A. *Nature* **376**, 19–20 (1995).
10. Sretavan, D., Shatz, C. & Stryker, M. *Nature* **336**, 468–471 (1988).

Watermarking by numbers

Jian Zhao

Is copyright dying in the digital age? Are we entering the chaos of an unregulated press and free printing, as happened during the French Revolution? We can copy photos, music and movies from the Internet, CDs, digital tapes and digital video disks, and easily distribute them to our friends, usually unconscious of guilt. But a new technology called digital watermarking¹⁻³ has arrived to handcuff pirates and deter illicit copying.

Like a traditional watermark on paper, which is only visible when held up to the light or chemically processed, a digital watermark is information hidden within a data set. It can only be detected by a computer program that has the same secret element as was used in embedding the watermark. Digital watermarking can be used to claim ownership, by embedding permanent author identifiers when the work is created, and then to track distribution, by adding a customer identifier when the work is copied. Other information such as user rights, contact addresses and key words can also be embedded, for a wide range of applications. And to remind the user of copyright, a public notice or logo can be embedded as a watermark in a publicly known location.

In addition to being imperceptible, the watermark must be hard to remove and tightly bound to the quality of the original data; that is, it must be resistant to common manipulations (such as compression, low-pass filtering, cropping, rotation and scaling) and to any attack that does not seriously degrade the value of the data. Obviously, a forger cannot afford to damage an image or a soundtrack in order to remove the watermark, and Cox *et al.*¹ point out that a watermark should be inserted in "perceptually significant" parts of a signal for that reason.

Because a good watermarking scheme is tolerant of some manipulations (such as format conversions), but sensible to the alteration of semantics (such as shifting or replacing an object in a photo), it provides a kind of authentication that is less strict than the 'digital signature' technique.

Digital watermarking is a form of steganography, the art of hiding secret messages within other, harmless messages. One ancient method was to tattoo a secret message on the scalp of a trusted slave's shaved head; once his hair had grown, he was dispatched to the destina-

tion and his head was shaved again. In this century, steganography was most famously practised during the Second World War, when Germany developed microdots and various invisible inks, and the US marines used the Navajo Indians' obscure native tongue for transmitting secret radio messages.

So how can we hide messages in digital information? The scheme of Cox *et al.*¹ is one example, based on a 'spread-spectrum' technique⁴: a narrow-band signal is transmitted over a large bandwidth, so that the spectral power density



Out of sight. The original image (left) was watermarked with the identification code 'Nature' using SysCoP. The result (centre) shows that the watermark is imperceptible — it is hidden in spectral noise. That image was then compressed, losing 80% of its information content (right), but it still contains the watermark.

is much lower and the signal just looks like noise. Our human visual and auditory system is not as precise as a good digital representation, and so it is highly tolerant of noise. That is why digital watermarking works: the watermark information is encoded into the original data as additional 'natural' noise, or by rearranging existing noise. Noise can be inserted not only into the spectrum, but also into other characteristics of the data — the contour or texture of images, for example.

In addition to the robustness and imperceptibility of the watermark, security is a major concern. To understand the possible forms of attack, we'll start by calling the original data D , the watermark W , and the watermarked data D_W . Then watermark insertion and extraction can be described as the following two equations: $D_W = D (+) W$, and $W = D_W (-) D$ where (+) and (-) are the operations of modulation and de-modulation.

An obvious counterfeiter's attack is over-watermarking — simply embedding a second watermark W' into the watermarked data D_W to produce $D_{W'}$. But this can be easily detected by the true owner, using the original data: with any workable watermarking scheme, W can be extracted from either modified data set, but W' cannot be constructed from the original, D . Moreover, uncontrolled over-

watermarking results in an accumulation of noise, and so might create visible or audible defects.

Recently, Craver *et al.*² have discovered another way to invalidate the real owner's claim. An attacker can use an invertible watermarking scheme to create a counterfeit original by 'subtracting' a new watermark W' from his copy of D_W , so $D_{W'} = D_W (-) W'$ (1) and, with luck, $D_{W'} = D (-) W'$ (2). In order to satisfy the second equation, the forger needs to know exactly where the first watermark was embedded. Then he can generate a second watermark in such a way that not only does D_W contain both W and W' , but D can be shown to contain W' , and so the real owner can't claim ownership. Of course, compulsory watermark registra-

tion and deposit would always clear up the confusion — the first to register must be the true owner.

Content-dependent watermarks will make it very difficult for the attacker to find such a W' (without access to D) satisfying equation (2). Furthermore, randomly selected watermark locations will lead the attacker to darkness and uncertainty: without knowing where W was embedded in D , he will be never sure that he has forged a W' that satisfies the equation².

Digital watermarking technology has already entered the commercial market, even though it has only existed in research labs for a short time, and both the technology and the theory are in their infancy. Soon it will be an important tool for establishing digital intellectual property rights — a problematic but vital aspect of the information age. □

Jian Zhao is at the Fraunhofer Center for Research in Computer Graphics, Providence, Rhode Island 02906, USA (e-mail: jzhao@crpg.edu).

1. Cox, I. J., Killian, J., Leighton, T. & Shamoon, T. Tech. Rep. 95-10 (NEC Research Institute, 1995).
2. Craver, S., Memon, N., Yeo, B. L. & Yeung, M. Tech. Rep. RC 20509 (IBM, 1996).
3. Berghel, H. & O'Gorman, L. *IEEE Computer* **29**, 101-103 (1996).
4. Dixon, R. C. *Spread Spectrum Systems with Commercial Applications* (Wiley, New York, 1994).

Genes from zebrafish screens

Nigel Holder and Andrew McMahon

LARGE-scale genetic screens have had a major impact on our understanding of invertebrate and plant development, yet until now there has been no systematic analysis of vertebrate embryogenesis. In a special issue of *Development* (Vol. 123, December 1996), Janni Nusslein-Volhard and Wolfgang Driever's groups report the identification of several thousand new mutants in the zebrafish *Danio rerio*; the mutations occur in hundreds of genes, leading to a range of developmental defects. These papers have been eagerly anticipated in the vertebrate-development

embryo and move to different regions of the body to form many tissues, including much of the head skeleton and the peripheral nervous system. Although the genes responsible for the formation of the crest may be present in *Drosophila*, they would need to be adapted for these new functions. So the comparative evolutionary approach can only go part of the way.

Up to now, genetics has been the most productive route for the identification of the major players in development. The concept is simple — mutate genes at

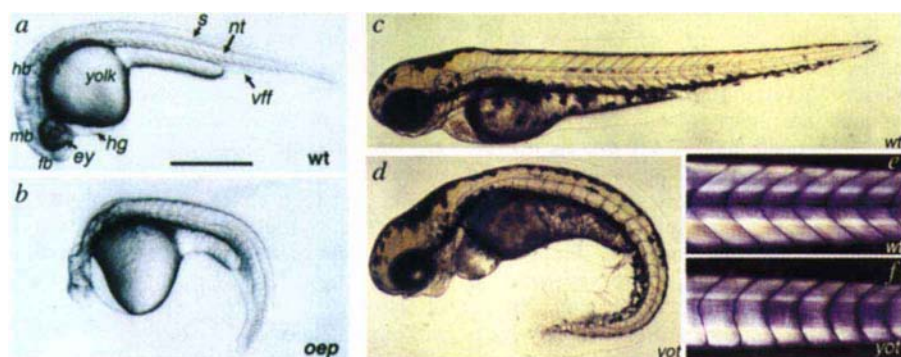
cost and organization for such a project would be enormous. Consequently, no systematic genetic screening programme has been carried out to search for mouse mutants that affect development, although many interesting spontaneous mutants have been identified.

Until recently, mouse genetics was limited to the analysis of phenotypes that can be easily scored in the neonate or adult. Now, a more powerful approach is the targeted mutation of genes that are likely to be of interest because of the known properties of the proteins they encode, their interesting expression patterns, or their relationship to regulatory genes in invertebrates. Four years ago, when vertebrate genetic screens comparable to those carried out on flies were being considered, only a handful of genes had been knocked out using homologous recombination in mouse embryonic stem cells. By contrast, in the past year, nearly 200 new mutations have been published. This technology has become essential to our progress in understanding many aspects of gene function, but the approach contains a considerable bias — you can only knock out what you have. So the investigator calls the shots, not chance.

The alternative is a small teleost, the zebrafish *D. rerio*, which was being groomed for success in the 1970s by a group of laboratories at the University of Oregon at Eugene. A virologist there, George Streisinger, wanted to analyse the development of the vertebrate visual system using genetics. He chose the zebrafish because it had all of the characteristics needed for genetic studies: it is easy to keep in the laboratory; it has (for a vertebrate) a reasonably short breeding cycle of 2–3 months; it develops very rapidly into a fry; and it is transparent through the early parts of its life (see figure).

Although *D. rerio* has fins instead of limbs, no lungs, and a primitive pronephric kidney, there are thought to be molecular parallels between the fin and limb development, and many of the genes that underlie pronephric development may also encode components of the complex metanephric kidney. Moreover, there are clear parallels between fish and mice in the expression of all of the crucial developmental-control genes, from the Hox complex to the Wnt family. Even such complex structures as the brain are much the same in basic design.

Crucially for Streisinger, zebrafish eggs are fertilized externally and then immediately dispersed, so individual embryos can be observed and scored for a specific trait. An alternative fish species that Streisinger could have selected was the Japanese medaka (*Oryzias latipes*), which has a history of genetic studies. But the mother keeps her brood of developing embryos in a sticky mass, tight to her cloaca, making it difficult to score embryos for a phenotype.



a, A 24-hour-old wild-type zebrafish embryo: ey, eye; fb, forebrain; hb, hindbrain; hg, hatching gland; mb, midbrain; nt, notochord; s, somite; vff, ventral fin fold. b, The mutant, *one-eyed pin-head* (*oep*), at the same stage as the wild-type embryo shown in a: *oep* lacks the head axial mesoderm and the ventral central nervous system. c, A wild-type larva at 72 hours of development. d, A *you-too* (*yot*) mutant larva at the same stage as the wild-type larva shown in c. The *yot* mutants comprise a small group that have a normal notochord but no myoseptum, and the somites form a U-shape, as shown in high-power views of wild-type (e) and *yot*-mutant (f) somites. These pictures are from the papers by Solnica-Krezel *et al.* from the Boston screen, and van Eeden *et al.* from the Tübingen screen, in the special issue of *Development*. (Figures courtesy of J. Nusslein-Volhard, P. Hafter and W. Driever.)

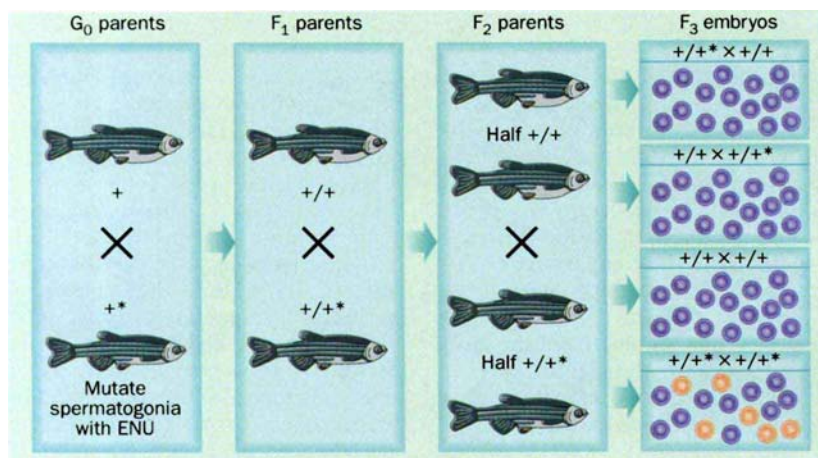
community, but how do the results match up to the high expectations, and what can we learn from them?

Remarkably, many of the genes that are important for the control of fly development are also crucial players in vertebrate, and by association human, development. This is lucky because it means that much of the identification and some of the functional analysis of key genes can be done in invertebrates such as *Drosophila* and *Caenorhabditis elegans*. Some of the similarities are amazing: for example, mutations in both the human *Pax6* gene and in *eyeless*, the *Drosophila* homologue, cause abnormal eye development. This maintenance of function occurs in spite of the overtly different manner in which *Drosophila* and human eyes develop.

Despite this kind of surprise — which is becoming a frequent event — there are clearly some developmental structures that are unique to vertebrates. For example, the neural crest is a group of migratory cells that begin on the back of the

random and look for abnormalities in particular biological processes. Interesting genes can be identified in this way, and new principles may even emerge. Testament to this are the screens performed by Nusslein-Volhard and Eric Wieschaus on *Drosophila*, in which they searched for (and found) many of the important genes that control early patterning events. The background to the screen that they performed is shown in the box (see over). But would such an approach work in a vertebrate?

The first challenge is to identify the animal to use. The mouse has the greatest history of genetic studies, and it is fairly well understood at the anatomical and physiological levels. So there would be a great advantage in a large collection of mouse mutants, not least because new models of human disorders are likely to arise. However, it is difficult to do a random-mutagenesis screen in the same way as with flies, because the embryos develop *in utero*, so mutant phenotypes cannot easily be observed. Moreover, the space,

The F₂ screen

THE F₂ screen was developed by Nusslein-Volhard and Wieschaus to identify new genes in *Drosophila*, and it has now been used in zebrafish. Adult male fish are treated with ethyl nitrosourea (ENU), which generates point mutations in DNA. The chemical concentration and treatment regimes mean that, on average, just over one mutation is introduced per genome. Males are treated because mutations in spermatogonia — the stem cells of the male germ line located in the testis — can be conveniently introduced into the gene pool in large numbers from

relatively few (40–50) fish. The resulting F₁ progeny are heterozygous for a unique set of mutagenized genes. F₂ families are then raised from single pairs of F₁ fish and, in any one family, 50 per cent of the siblings will be heterozygous for the same mutation. F₂ siblings are then crossed, and two siblings that are heterozygous for the same mutation will be crossed with each other, on average, in 25 per cent of cases. A quarter of the F₃ embryos created from these crosses will be homozygous for a particular mutation, and these are scored visually. □

So the zebrafish got the nod. It is one of the sad ironies of research that, having had the initial vision and done much of the hard work, Streisinger died in 1984 before he could bring his screens to fruition. The survival of the zebrafish as a vertebrate developmental system is largely due to Charles Kimmel and his colleagues in Oregon. They took up the mantle of Streisinger and put zebrafish developmental genetics on a firm footing, from which the Driever and Nusslein-Volhard groups took off.

In parallel, the two groups set out to saturate the zebrafish genome with mutations and to identify all of the major developmental-control genes. The scale of the enterprise was impressive — Nusslein-Volhard's *fischhaus* in Tübingen contained some 6,000 fish tanks, and Driever's group in Boston had almost as many. Following the initial F₂ screen (see box), complementation crosses were done to identify how many genetic loci had been affected. The screens were designed to reveal mutations that had a zygotic effect in a range of developmental events, such as establishment of the embryonic axes; patterning of the germ layers (ectoderm, mesoderm and endoderm); formation and morphogenesis of organs such as the heart, brain and sense organs; and

generation and differentiation of neural crest. Mutants were identified by visual inspection of the largely transparent embryos and larvae at key stages during the first five days of their life.

More than 1,000 mutations were obtained in both screens. From the Boston screen, 577 mutations are reported in the special issue of *Development*, and complementation analysis has so far confirmed a total of 220 genes. The Tübingen screen reports 1,163 mutations, representing 369 genes. Many of the mutations give rise to fascinating phenotypes, affecting the development of various organ systems and the formation of the early embryonic axes. There is no doubt that these mutants represent an impressive outcome and an unparalleled resource for vertebrate developmental biologists.

Despite their obvious success, the screens did not achieve everything their originators might have initially hoped for — they were a learning process for all concerned. Because the groups only searched for obvious visible phenotypes, many potentially interesting mutations may have gone undetected or have been discarded prematurely. Also, it is not always clear what phenotype to expect from a mutation in a key gene. The original Nusslein-Volhard and Wieschaus screen in

Drosophila indicated that failure to make important cell-fate choices often resulted in cell death. And extensive cell death was observed in many of the discarded mutants. So were interesting mutants lost amongst those that affected the less interesting housekeeping functions? It is clear from fly and worm genetics that the genes that are identified are defined by the type of screen performed.

An indication that the zebrafish screens did not reach saturation is the rate of generation of alleles for a particular gene. Although more than 10 alleles were found for some genes, only one allele could be found for many others. It is likely that, as in *Drosophila*, many of the main early control points are governed by maternally encoded information, and that a maternal-effect screen will be needed to examine some of the most interesting and elusive aspects of vertebrate embryogenesis. Other small-scale screens that have a particular developmental process in mind will also probably be successful in identifying new loci. One such ingenious screen, performed in Friedrich Bonhoeffer's laboratory, focused on the retinal-tectal connection in the formation of the visual system. They used dye-tracing from the retina to the tectum to uncover a number of hitherto unknown genes that alter axonal projections.

It is one thing to find the genes by mutation, but how can they be cloned? The answers to this persistent question were discussed at a meeting at Cold Spring Harbor (April 24–26 1996, "Zebrafish Development and Genetics"). A detailed genetic map complete with random markers spaced at approximately 3-centimorgan intervals is now available, and the resolution is likely to improve in the near future. The map has already been used to identify a cloned gene as the culprit for a specific mutation, and it should soon be used for the first time to positionally clone a gene. A yeast artificial chromosome (YAC) library is also available, which may be used to rescue the phenotype following injection into the fertilized egg. The gene of interest could then be identified by deletion analysis with the rescuing YAC. Somatic-cell hybrid panels and radiation hybrids are available, and there are definite regions of synteny (conservation of chromosomal structure) between zebrafish, pufferfish and human chromosomes. It shouldn't be long before we are able to clone this plethora of interesting genes — Streisinger would have been pleased. □

Nigel Holder is in the Developmental Biology Research Centre, Randall Institute, King's College, 26–29 Drury Lane, London WC2B 5RL, UK. Andrew McMahon is in the Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

Train signals for electric fish

Len Maler

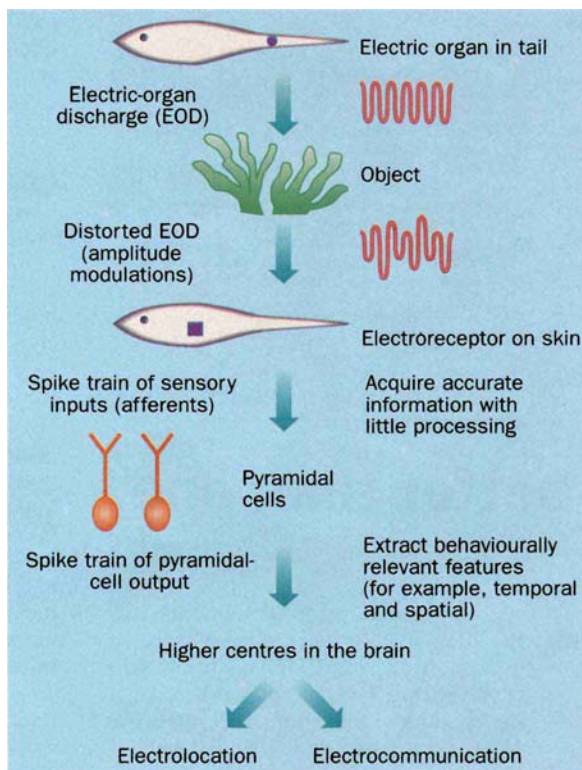
ANIMAL nervous systems are subjected to an incoming stream of sensory data from which they extract specific features. It is not usually possible to quantify these features, but on page 564 of this issue Gabbiani *et al.*¹ describe a new method for doing just this. Using the weakly electric fish *Eigenmannia viriscens*, they show that electrosensory neurons can encode temporal features of their sensory input. This quantitative analysis provides the key to understanding how cellular properties such as dendritic integration and spike bursting contribute to the detection of sensory features.

Electric fish use an electric organ that is derived from muscles or motor nerves in their tails to produce an oscillating electric-organ discharge (EOD) that helps them to detect their environment. The EOD stimulates electroreceptors in the skin², and any object with electrical properties that are different from those of water will distort the field and be detected (see figure).

Gabbiani and colleagues have used a South American species that generates a quasi-sinusoidal high-frequency EOD. Most electroreceptors in this species encode the amplitude of the EOD by the probability, per EOD cycle, that the electroreceptor will be activated (release its neurotransmitter), and cause the peripheral nerve fibre that innervates it to discharge an action potential. In this way, the amplitude of the EOD is translated into a pattern of action potentials (known as the spike train), which travels up the peripheral nerve and into the brain. Electroreceptor afferent nerves enter the brain and make synaptic contacts in an enlarged region known as the electrosensory lateral-line lobe. This is found in the hindbrain, and it is analogous to auditory regions in the mammalian hindbrain. Pyramidal cells in the electrosensory lateral-line lobe signal spatial and temporal variations in the amplitude of the EOD. This information is projected to higher centres in the brain that control locomotion (electrolocation) and communication (electrocommunication).

An earlier study from the same group³ used a spike-train decoding method⁴ to show that electroreceptors can accurately follow random EOD amplitude modulations (AMs). The technique was based on

the determination of a linear filter which, when combined with a spike train of sensory inputs (afferents), produced an optimal estimate (in the least-mean-square-error sense) of the stimulus driving that afferent. Electroreceptor afferents could encode up to 75% of the information in the AM stimulus. But surprisingly, when the same technique was applied to pyra-



The electric fish *Eigenmannia viriscens* generates a quasi-sinusoidal electric-organ discharge (EOD), the local amplitude of which varies as the fish moves across objects that have a different resistivity from that of water. The EOD stimulates electroreceptors on the skin of the fish, and the electroreceptor discharge frequency is determined by the amplitude of the local EOD. The electroreceptors project to pyramidal cells in a region of the central nervous system called the electrosensory lateral-line lobe. Gabbiani *et al.*¹ show that neurons in the electrosensory lateral-line lobe do not follow the amplitude variations of the EOD: instead they encode temporal features (or vectors) that are interpreted by higher centres in the brain that control locomotion and communication.

midal cells, they were found to encode less than 30% of the total stimulus¹. So Gabbiani *et al.* have now tried to ascertain whether this difference could be explained by temporal features in the stimulus driving the spiking of pyramidal cells.

The key assumption of their analysis is that an optimal sensory stimulus will maximize the signal-to-noise ratio of the spike-train response by the pyramidal cells. The first step in the computation is to determine the mean stimulus that precedes spiking by the pyramidal cells, and the

mean stimulus that leads to a null response. The difference between these stimuli represents the average stimulus that precedes spike firing — that is, the 'average stimulus vector'. The next step is to establish the input that optimally predicts when a pyramidal cell will spike. The authors define this as the 'feature vector', and they propose that the pyramidal cells will most easily distinguish this particular vector from background noise.

To optimize the signal-to-noise ratio in this way, a feature vector must satisfy two conditions: it must be maximally similar to the average stimulus vector defined above and, at the same time, it must minimize the variability (statistical variance) of the pyramidal-cell spike train. The latter point can be intuitively appreciated by considering that it is difficult to detect a signal that is buried in a noise of equal power. So although a signal may cause a pyramidal cell to spike, the signal will not be optimal if the response by the pyramidal cells varies from the presentation of one stimulus to the next. By contrast, the feature vector will evoke a strong and consistent response from the pyramidal cells.

This technique has revealed that not only do pyramidal cells respond to fluctuations in the amplitude of the EOD, but that they also temporally integrate the stimuli for about 300 ms before they spike. So although the spike-decoding method is well suited to the analysis of primary sensory-afferent spike trains, the new feature-vector technique may be a superior tool for the study of central sensory pathways, because neurons at higher levels of the nervous system do not simply track the input from the peripheral nervous system; they also extract particular features from this input. The new technique has advantages over traditional methods used by sensory physiologists (for example, the use of artificial stimuli such as step changes in an input²), and it is complementary to the neuroethological approach, which analyses sensory input in terms of the species-specific behaviours it evokes⁵. But these methods do not allow the extent of the temporal integration that is done by pyramidal cells to be detected^{2,5}. Although temporal coding is generally recognized as being important for specialized neurons in the auditory and electrosensory brain stem⁶, the new study implies that coding for temporal features may, in fact, be widespread in neurons that seem to be simply signalling changes in amplitude.

Gabbiani *et al.*¹ also describe two modes of pyramidal cell discharge — individual

spikes versus bursts of spikes. Spike bursts are often described in mammalian cortical neurons⁷, but it is not known why they occur. Remarkably, the spike bursts that signal the occurrence of the feature vector have the highest signal-to-noise ratio compared with primary afferents. The feature-vector method is based on a statistical model of nonlinear stimulus discrimination: that is, a linear summation of inputs followed by a thresholding step. A simple model would predict that the stimulus amplitude modulation that is encoded by electroreceptors would be present as the summed membrane potential at the cell bodies of pyramidal cells, and that the threshold would be spike generation near the cell body. But it seems that this model is incorrect, because the authors have found that the summed potential at the cell body does not represent the stimulus better than the spike train.

Turner *et al.*⁸ showed that spike bursting by pyramidal cells in the electrosensory lateral-line lobe is caused by action potentials at the pyramidal cell body that are generated by currents arising from dendritic spikes. So Gabbiani *et al.* have proposed that temporal feature extraction

might require spike bursts generated by a complex interaction between currents at the pyramidal cell body and dendrites. This hypothesis may be generally applicable — a recent modelling study⁹ has indicated that the mechanism by which spike bursts are generated, through interactions between dendrites and the cell bodies, might also operate in mammalian cortical neurons. □

Len Maler is in the Department of Anatomy and Neurobiology, University of Ottawa, 451 Smyth, Ottawa, Ontario, Canada K1H 8M5 (e-mail: lmaler@aix1.uottawa.ca).

1. Gabbiani, F., Metzner, W., Wessel, R. & Koch, C. *Nature* **384**, 564–567 (1996).
2. Bullock, T. H. & Heiligenberg, W. *Electroreception* (Wiley, New York, 1986).
3. Wessel, R., Koch, C. & Gabbiani, F. *J. Neurophysiol.* **75**, 2280–2293 (1996).
4. Bialek, W., De Ruyter, R. R., Van Stevenick, R. & Warland, D. *Science* **252**, 1854–1857 (1991).
5. Heiligenberg, W. *Neural Nets in Electric Fish* (MIT Press, Cambridge, 1991).
6. Carr, C. E. *Annu. Rev. Neurosci.* **16**, 223–243 (1993).
7. Franceschetti, S., Guatteo, E., Panzica, F., Sancini, G. E. W. & Avanzini, G. *Brain Res.* **696**, 127–139 (1995).
8. Turner, R. W., Maler, L., Deerinck, T., Levinson, S. R. & Ellisman, M. H. *J. Neurosci.* **14**, 6453–6471 (1994).
9. Mainen, Z. F. & Sejnowski, T. J. *Nature* **382**, 363–366 (1996).

IMMUNOLOGY

Two faces are better than one

Bernard Malissen

T LYMPHOCYTES sense the infection of a cell by recognizing viral or bacterial peptides that are displayed at the cell surface in association with major histocompatibility complex (MHC) class I molecules. The T lymphocytes scan the surfaces of cells for peptide/MHC complexes that have the capacity to bind to their clonally variable T-cell antigen receptors (TCRs). The many specific TCRs on a T cell can then interact, one after the other, with a single peptide/MHC complex expressed on a target cell, and this is enough to elicit signalling within this T cell^{1,2}. To succeed in such a formidable task, the TCR requires the help of other surface molecules, including the accessory molecule CD8. And on page 577 of this issue, Garcia *et al.*³ describe studies that help us to understand more about the role of CD8 in immune recognition.

CD8 is expressed at the surface of T cells, and it consists of two disulphide-linked chains that form $\alpha\alpha$ homodimers and $\alpha\beta$ heterodimers. CD8 α has an amino-terminal extracellular domain, with a fold that is typical of immunoglobulin variable domains. This is followed by a 45-amino-acid connecting peptide that precedes a 25-amino-acid transmembrane segment, and a 28-amino-acid cytoplasmic tail⁴. The CD8 β -chain shows a similar overall organization; however, CD8 $\beta\beta$ dimers cannot be formed, and CD8 β

requires the presence of CD8 α to reach the T-cell surface.

Some time ago CD8 was found to be expressed on T cells capable of recognizing peptides that are associated with MHC class I molecules, and it later became clear that CD8 recognizes the MHC class I molecules themselves. Using an intercellular adhesion assay, Salter *et al.* were among the first to show that CD8 $\alpha\alpha$ dimers can bind to a solvent-exposed loop within the third domain of class I MHC molecules — regardless of the presence of the TCR (ref. 5). Importantly, these data indicated that the TCR and CD8 bind to distinct sites on the class I MHC. They further implied that a single MHC class I molecule can be simultaneously bound by a TCR and by CD8 (see figure, over).

Because CD8 also associates with the intracellular tyrosine kinase p56^{lck}, CD8 could potentially mediate intercellular adhesion and promote or enhance TCR signalling. After the TCR and CD8 have engaged the same MHC molecule through their respective extracellular domains, their cytoplasmic segments will be juxtaposed so that the TCR transducing subunits or their associated effectors (for example, ZAP-70) can be phosphorylated by p56^{lck}, contributing to the successful progression of the activation cascade.

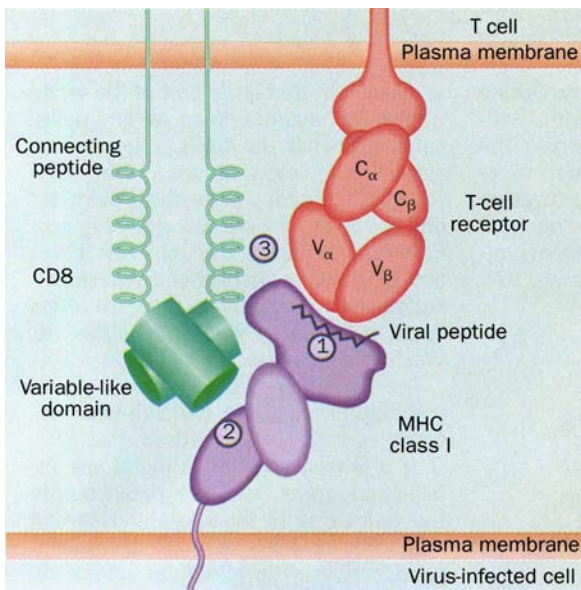
Pioneering studies by Allen, Sette and

colleagues (reviewed in ref. 6) established that the TCR does not function as a binary (on-off) switch. They suggested instead that small differences in the affinity of the TCR for its ligand, the peptide/MHC complex, translate into markedly different signalling outcomes. To explore whether the half-life of TCR-peptide/MHC complexes governs the progression into the T-cell activation programme, over the past few years many different strategies have been used to produce soluble forms of these components and to determine their binding kinetics. The TCR dissociation constant (K_d) for peptide/MHC complexes ranges from 10^{-4} to 10^{-7} M, confirming the view that the peptide/MHC ligands that cause partial T-cell activation tend to dissociate more rapidly than those that trigger full activation⁶.

But none of these measurements took into account the contribution of the CD8 molecule. To do this, Garcia *et al.*³ have used surface plasmon resonance to monitor the effect of CD8 on TCR-peptide/MHC interactions at 25 °C. They first showed that CD8 extracellular domains bind to various immobilized peptide/MHC complexes with K_d values ranging from 11 to 39 μ M and corresponding to a half-life for the ternary complex of between 14 and 16 seconds. Neither the polymorphic residues found in the MHC peptide-binding groove nor the primary structure of the peptide influenced the parameters of the CD8-peptide/MHC interaction. Interestingly, the binding of CD8 $\alpha\alpha$ homodimers and CD8 $\alpha\beta$ heterodimers to MHC class I molecules is similar. So the functional observation that the heterodimers are more effective than homodimers in enhancing T-cell responses probably reflects the different efficiencies with which their respective cytoplasmic segments interact with the p56^{lck} kinase⁷.

The authors next showed that in the absence of CD8, the peptide/MHC complex rapidly dissociates from the TCR. On adding CD8 $\alpha\beta$ dimers, they found that the affinity of a TCR for its specific ligand is enhanced, due to a reduction in the off-rate. The contribution of CD8 was even more marked in the case of a second TCR, which had a lower affinity for its peptide/MHC ligand. By showing that successful TCR-peptide/MHC-CD8 interactions can occur in a non-cellular environment — eliminating the possibility of the adventitious participation of other cell-surface molecules — Garcia *et al.* extend the contributions made by Luescher *et al.*⁸, who originally pointed out that CD8 enhances the lifetime of the interaction between a TCR and a peptide/MHC complex. Moreover, the new results provide a quantitative background to the observation that alterations in the levels of CD8 can influence the fine specificity of the TCR, so that an antigenic peptide that acts as an antagonist can even become a weak agonist⁹.

Gel-filtration chromatography and



Interaction between CD8 and a T-cell receptor (TCR) bound to a class I major histocompatibility complex (MHC) molecule loaded with a viral peptide. The extracellular part of CD8 consists of an immunoglobulin variable-like domain and a peptide that connects it to the intracellular part of the molecule. The connecting peptide contains many proline residues and O-linked carbohydrates, both of which cause it to adopt an extended structure that allows the CD8 variable-like domain to reach out to the membrane-proximal domain of a TCR-engaged MHC class I molecule¹². The TCR docks with the MHC peptide-binding groove at site 1; however, CD8 binds primarily to a surface involving the third MHC domain (site 2), as well as to the base of the peptide-binding groove of the MHC class I molecule^{5,13}. Garcia *et al.*³ suggest that the extracellular domain of CD8 (residues 1–156) interacts physically with the TCR (site 3).

native gel electrophoresis allowed Garcia *et al.* to confirm the existence of labile TCR-peptide/MHC-CD8 complexes. They also made the unexpected observation that CD8-mediated enhancement of TCR-peptide/MHC association does not require stoichiometric molecular ratios. In other words, the contribution of CD8 to the TCR-peptide/MHC interaction does not seem to rely on its persistent binding. It is tempting to speculate that to achieve a correct fit with the MHC peptide-binding groove, the TCR needs the transient assistance of CD8. So CD8 could constitute a kind of molecular chaperone, increasing the number of TCR-peptide/MHC complexes.

It is not clear to what extent TCR complementarity-determining regions change their conformation on forming a complex with peptide/MHC ligands¹⁰, and no TCR structure has yet been structurally resolved before and after forming this complex. But it has recently been proposed that the third complementarity-determining region of a TCR β -chain undergoes a large conformational change on binding to its cognate peptide/MHC complex (D. Housset *et al.*, manuscript in preparation). By transiently interlocking the TCR and the peptide/MHC complex, CD8 may allow the loop of the complementarity-determining region to adopt configurations that allow

the TCR to interact properly with the surface of the MHC peptide-binding groove. Consequently, CD8 may affect the binding of some TCRs more than others. Further experiments, taking into account the fact that interactions between the TCR and CD8 are normally constrained by diffusion within the plane of the T-cell membrane, are clearly required to deduce the functional implications of the TCR-peptide/MHC-CD8 interaction.

Finally, the data of Garcia *et al.*³ reinforce the view that CD8 constitutes an 'intrinsic component' of the TCR rather than a plain accessory molecule. What physiological benefit might result from the split of the MHC class I binding interface into two surfaces harboured by distinct transmembrane molecules? Ligand-induced dimerization is a common means by which receptors that have, or are associated with, a tyrosine-kinase activity transmit their signals to the inside of a cell. Such an effect has been elegantly demonstrated for

the growth-hormone-receptor complex¹¹ — a single molecule of growth hormone can bind to two identical receptor molecules, which together may trigger downstream phosphorylation events. The peptide/MHC class I molecules seem to behave like the growth hormone in that they use two different sites (sites 1 and 2 in the figure) to bring together the TCR and CD8. Only when these two components are brought together through binding to the same ligand will signal transduction occur optimally. The segmentation of the peptide/MHC class I binding site into two distinct polypeptides may, therefore, lie at the heart of the dimerization-activated switch that is used by the TCR. □

Bernard Malissen is at the Centre d'Immunologie, INSERM-CNRS de Marseille-Luminy, Case 906, 13288 Marseille, Cedex 9, France.

1. Sykulev, Y. *et al.* *Immunity* **4**, 565–571 (1996).
2. Valitutti, S. *et al.* *Nature* **273**, 148–151 (1995).
3. Garcia, K. C. *et al.* *Nature* **384**, 577–581 (1996).
4. Leahy, D. J. *et al.* *Cell* **68**, 1145–1162 (1992).
5. Salter, R. D. *et al.* *Nature* **345**, 41–46 (1990).
6. Sloan-Lancaster, J. & Allen, P. *Annu. Rev. Immunol.* **14**, 1–27 (1996).
7. Irie, H. Y. *et al.* *J. Exp. Med.* **181**, 1267–1273 (1995).
8. Luescher, I. F. *et al.* *Nature* **373**, 353–356 (1995).
9. Jameson, S. C. *et al.* *Nature* **369**, 750–752 (1994).
10. Garboczi, D. N. *et al.* *Nature* **384**, 134–141 (1996).
11. Wells, J. A. *Proc. Natl Acad. Sci. USA* **93**, 1–6 (1996).
12. Classon, B. J. *et al.* *Int. Immunol.* **4**, 215–225 (1992).
13. Sun, J. *et al.* *J. Exp. Med.* **182**, 1275–1280 (1995).

Inside stories

LAST week, Daedalus outlined a way of inverting living tissue. He reckoned that if it were treated with heavy water, the outer cell membranes would tend to warp concave. Ultrasonic agitation could then turn the whole system inside-out, so to speak. The cell interiors would merge into a continuous medium, within which closed cell membranes would now hold isolated drops of intercellular fluid. Many emulsions can be inverted this way.

Inverted tissue should continue to flourish, at least for a while. The cellular mass will continue to metabolize by exchange across the cell membranes with the intercellular medium. Even so, the obvious applications are destructive. Daedalus began to think of dosing a patient with heavy water, tuning an ultrasonic beam into some resonant mode of an invading bacterium, and focusing it into an accessible vein. As the patient's blood flowed through the vein, the passing bacteria would be turned inside-out, damaging them fatally.

Sadly, single-celled organisms usually have tough cell walls, not easily inverted. But the technique should work well on bigger parasites, such as those of malaria and similar diseases. They tend to hide from the body's immune system by expressing misleading antigens on their outer layers. Once inverted, their true internal identity would be revealed; they would be speedily mopped up. Even partial inversion of their internal cells should soon kill them.

Inversion might even be a useful way of limiting damage to the patient's own tissues. A precisely defined abnormal region could be inverted by a tightly focused ultrasonic beam, giving a cellular continuum which could be treated with powerful drugs that cannot easily cross cell membranes. On later re-inversion, the cells would be randomly reassembled. Some would form around two nuclei or other organelles in the continuum; others would find themselves with none. Pathogens, a minority in the mass, could never reassemble correctly. The result would be a sort of stable scar tissue. In fairly homogeneous organs, such as the brain, the treated region might even continue to function, though local obsessions, phobias, and so on, would be usefully erased.

Cellular inversion might also be a splendid form of fast cooking. An inverted joint of meat, say, would be wonderfully digestible. Its nutritive content would no longer be trapped within its cell membranes. Any bacteria in it would be turned inside-out and destroyed. Lovers of truly rare steak could eat it with perfect confidence.

David Jones

Abdus Salam (1926–96)

AMONG the great intellectual achievements of this century is the theory of elementary particles that has come to be known as the Standard Model. It represents the cumulative effort of many profound and imaginative thinkers whose common motivation was to find some glimmer of understanding as to what the physical world is made of and how it works. This kind of endeavour is very much in the European tradition, and much of the work was carried out in the wealthy universities of Europe and North America. But among the creators of this intellectual system there are some from other, less wealthy parts of the world. One such is Abdus Salam, who died on 21 November. Salam made a major contribution to the creation of the Standard Model. For this work he shared the Nobel prize with Steven Weinberg and Sheldon Glashow in 1979.

The Standard Model is the best existing mathematical description of the physical phenomena that take place on subnuclear scales. Its predictions have been confirmed to a high degree of accuracy. The main element in the construction of the model is a type of field theory known as Yang–Mills theory. In the 1950s Salam, together with J. C. Ward, was one of the first to appreciate the significance of Yang–Mills theory in the description of weak nuclear forces.

Then in 1962 an important theorem, due originally to Goldstone, was proved by Salam, Weinberg and Goldstone. Its extension and application to models of the Yang–Mills type by Higgs, Kibble and others led to an understanding of how the Yang–Mills vector particles can become massive. This was the last ingredient needed for the construction of a unified theory of weak and electromagnetic interactions, and the electroweak theory of Weinberg and Salam soon followed.

Abdus Salam was born in 1926 into an ordinary family in the village of Jhang in Punjab, in what was then British India. His father was a deeply religious person and placed great emphasis on education; and he was the single most formative influence on Salam during his education in Pakistan. Salam was educated at the government college in Lahore, at St John's College, Cambridge, and at the Cavendish Laboratory, where he obtained his PhD in 1952. In 1957 he was invited by Blackett to occupy the chair of theoretical physics and to create a theoretical particle physics group at Imperial College, London. In 1959 he was elected fellow of the Royal Society.

A guiding principle during his entire

active life was that of symmetry. During the early 1960s, together with Gell-Mann and others, he emphasized the group-theoretical classification of hadrons. Salam's contribution, together with John Strathdee, to a rather abstract form of symmetry known as supersymmetry is well known. In 1974 and 1975



they invented a mathematical framework of supersymmetry known as super-space. As an aside, it may not be well known, even among the experts, that several terms in common use by physicists, such as electroweak theory, astroparticle physics and supersymmetry, were invented by Abdus Salam.

Also in the mid-1970s, Salam and Robert Delbourgo were the first to formulate a possible violation of the equivalence principle in general relativity, due to the quantum effects of chiral fermions. This work influenced the development of gravitational instantons by Stephen Hawking and others. Salam and Jogesh Pati were among the first group of theoreticians to propose the idea of a grand unified theory encompassing electroweak and the strong nuclear forces, and that in such models the proton may be unstable.

The idea of creating an international centre where scientists of all nationalities and political creeds could interact was rooted in Salam's experience of being isolated from world science in Pakistan. Due mainly to his efforts, the International Centre for Theoretical Physics (ICTP) at Trieste, Italy, was established in 1964 under the aegis of the International Atomic Energy Agency and with the support of the Italian government. It was directed by Abdus Salam from 1964 until December 1993.

Even if he had not been such a creative and influential scientist, Salam would have been remembered for the creation of the ICTP and for his charismatic leadership. The centre has had a significant impact on the development

of science in that large part of the world where, perhaps due to social and political conditions, the basic sciences are almost completely ignored. The ICTP is, and has been, a meeting point for physicists of all cultures and religions. For example, during the Cold War it was one of the few institutions where scientists from the Soviet bloc countries could meet and collaborate with their Western colleagues.

The ICTP has overcome many seemingly insurmountable difficulties to become as well established as it is now. Only a person of Salam's moral and intellectual power, fired by a dedication to the well-being of the neglected part of humanity, could have made this enterprise survive. His relentless efforts on the political front are legendary. He has talked with some of the most important political leaders of his time, from John F. Kennedy and Zhou En-lai to François Mitterrand and Margaret Thatcher.

Abdus Salam was elected to membership of prestigious societies in 24 countries, including the US National Academy of Sciences, the Royal Swedish Academy of Sciences and the USSR Academy of Sciences; he was made an honorary KBE in 1989 and was an honorary member of seven other important national orders in four continents; he received 45 Doctor Honoris Causa awards in 28 countries, and nine medals for his contributions towards peace and the promotion of international collaboration, including the Atoms for Peace Award. Salam also founded the Third World Academy of Sciences and was its president until 1994.

During the last years of his life Salam suffered from a neurological disorder, progressive supranuclear palsy. This was perhaps the only significant battle, among so many, that he fought but did not win. He bore the illness with grace and tranquillity. He accepted with equal humility the great gifts that enabled him to act so effectively in science and for the Third World, and to face the vicissitudes of illness.

It was difficult not to be impressed by the forceful but deeply humane personality of Abdus Salam. He had a genuine respect for others and a very intelligent sense of humour. What he generously gave to others were his time and inspiration, perhaps the most valuable gifts that such a singular person could offer to his fellow human beings.

Seif Randjbar-Daemi

Seif Randjbar-Daemi is at the International Centre for Theoretical Physics, 34014 Trieste, Italy.

Evolutionary handicap for turtles

SIR — Sea turtle populations are facing an increasing threat from anthropogenic developments along shores. The long-term effects of the elimination of sea turtle nesting beaches have been unpredictable because of the lack of information on the precision of natal homing behaviour and gene flow between nesting sites. For the endangered sea turtle *Caretta caretta* we quantified genetic diversity in nuclear and mitochondrial DNA within and between adjacent nesting sites in the eastern Mediter-

anean Sea. The results demonstrate a low male-mediated gene flow between nesting sites and the presence of genetic substructuring due to the precise natal homing behaviour of females. This behaviour probably represents an evolutionary constraint and handicap for sea turtles to react to sudden habitat alterations.

Beginning at an age of 15–30 years, female sea turtles migrate hundreds or thousands of kilometres from their feeding to their nesting habitats¹. The

magnetic compass orientation² may guide females to the coast where they were born, and several studies have demonstrated natal homing behaviour at coastal areas separated by 1,000 km or more^{3–5}. For conservation planning, however, more information on the precision of natal homing is highly desirable. Furthermore, a knowledge of genetic variability between closely neighbouring nesting sites should help to reconstruct immigration events.

For one endangered species, the loggerhead turtle *C. caretta*, we obtained genetic data from geographically adjacent nesting sites in the eastern Mediterranean Sea, where building operations along the sea shore have become a steadily increasing threat. In 1993 and 1994 we collected tissue samples from 81 *C. caretta* hatchlings, which had been found dead at 10 different nesting sites in the eastern Mediterranean. From these samples we estimated the degree of the female-only mediated genetic differentiation by sequencing a part of the mitochondrial control region of 30 representatives from 10 nesting sites (Fig. 1).

In addition, we examined the overall (male- and female-mediated) genetic differentiation at the level of the nuclear genome by means of randomly amplified polymorphic DNA (RAPD^{6,7}) analysis between five closely neighbouring Turkish nesting sites (Fig. 2). Both the nuclear and mitochondrial DNA (mtDNA) data provide compelling evidence for a genetic separation not only between coastal areas (for example, Greece and Turkey), but also of colonies from adjacent nesting sites.

Precise natal homing is the straightforward explanation for these findings. The coexistence of two mtDNA haplotypes, I and III, in the Mediterranean, as well as their presence in the Atlantic (identical to A1cr and A3cr, respectively⁸) suggests independent immigrations of at least two different matrilineal lineages from the Atlantic into the Mediterranean. These may have occurred — at the earliest — after the last glacial period⁹. Because of the short post-glacial history of the Mediterranean loggerhead colonies, one may not expect to find fixed mtDNA haplotypes yet, and the gradual decline of haplotype III from the west to the east of Turkey (Fig. 1) probably represents an intermediate stage to fixation. Since post-glacial immigration, the natal homing behaviour must have caused demographic isolation between different populations within the Mediterranean, leading to the genetic differentiation in the nuclear and mitochondrial genome.

Despite the paucity of information on the reproductive behaviour of males

FIG. 1 DNA sequence analysis of the mitochondrial control region of *C. caretta* in the eastern Mediterranean Sea. We designed primers for polymerase chain reaction (PCR) amplification according to ref. 4. After sequencing 518 base pairs from 30 individuals, we found 11 different haplotypes. One widespread haplotype (I) was shared by individuals from Turkey ($n=8$), Cyprus ($n=2$) and Zakynthos ($n=4$). The other haplotypes (ht) (which will be discussed in detail elsewhere) were limited to geographically well-defined areas: the coasts of Greece ($n_{ht}=5$), Turkey ($n_{ht}=3$), Cyprus ($n_{ht}=1$) and Crete ($n_{ht}=1$). Further evidence for genetic separation of colonies from adjacent sites derives from distribution patterns of haplotype III, which was observed only in Turkey. This haplotype could be identified by restriction of the amplified fragment with Alu I, as haplotype III was the only one lacking an Alu I restriction site. Examining 81 individuals, we found that the frequency (F) of haplotype 'III Alu I⁻' significantly decreased from west to east along the Turkish coast over short geographical distances ($P < 0.001$, Jonckheere test of monotonous trend). West, Dalyan and Fethiye ($n=24$, $F=42\%$); Mid, Patara and Kumluca ($n=15$, $F=33\%$); East, Kizilot and Anamur ($n=19$, $F=16\%$). Haplotype III is absent in Cyprus ($n=3$) and Greece ($n=20$). The complete sequence and variable sites have been deposited in GenBank (accession no. U72747).

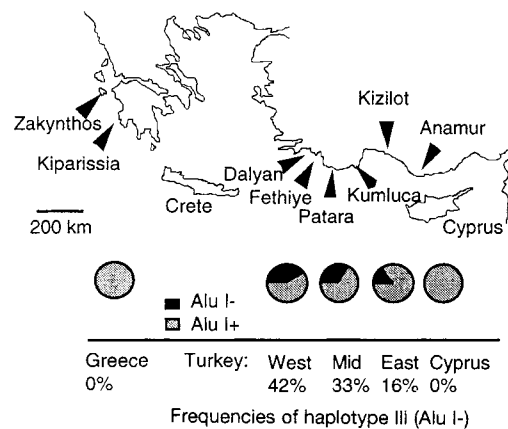
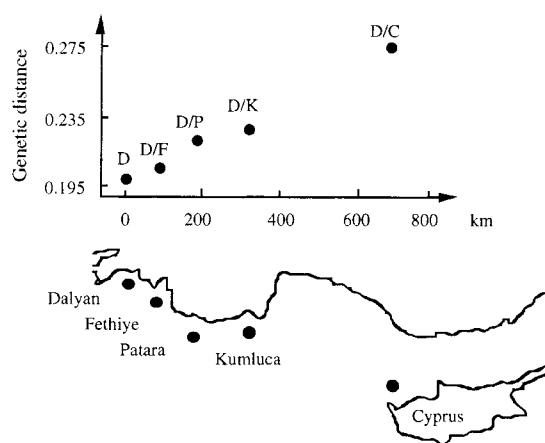


FIG. 2 Results of RAPD analysis for 40 *C. caretta* specimens from geographically adjacent nesting sites along the Turkish coast. D, Dalyan ($n=9$); F, Fethiye ($n=13$); P, Patara ($n=7$); K, Kumluca ($n=8$); C, Cyprus ($n=3$). We carried out RAPD-PCR¹¹ using five random 10-base oligonucleotide primers (AB4, M4, M9, M12, P2; Operon Tech.) generating 152 genetic markers, 71% of which were polymorphic. Estimates of genetic distances within and among colonies from RAPD markers (dissimilarity coefficients) revealed a strong correlation between increasing genetic and geographical distance (Mantel test, $r=0.61$, $P < 0.003$ over 9,999 permutations). The value for the reference colony Dalyan represents the genetic distance between individuals within this colony; all other values represent the mean genetic distance to the reference colony. Dissimilarity coefficients derive from pairwise comparisons of RAPD fingerprint profiles transformed into a phenotypic distance matrix. An 'analysis of molecular variance' (AMOVA statistics¹²) suggests a considerable amount of genetic variance accounted for by differences among colonies ($\phi_{st} = 0.106$; $P < 0.005$; 9,999 permutations) indicating genetic substructuring and limited gene flow between colonies.



and the corresponding gene flow¹⁰, the nuclear DNA data nonetheless suggest that a possible male-mediated gene flow between colonies is insufficient to prevent a genetic substructuring even of closely neighbouring colonies.

The consequences for conservation management are obvious. The pre-programmed female reproductive behaviour makes it unlikely that the loss of a breeding habitat (because of human building operations, for example) can be compensated for by emigration to other colonies; that is, the loss of nesting sites is accompanied by the loss of specific genotypes. Thus, to preserve the genetic diversity of the *Caretta* metapopulation one needs to preserve individual nesting sites (attempts to transfer freshly deposited eggs between sites — based on the hope that hatchlings would accept a new hatching site as natal — are controversial). In addition, the described haplotypes may serve as genetic tags to identify the colony membership of individuals offshore or illegally commercialized, and also to complete our knowledge of sea turtle migratory behaviour offshore¹¹.

Werner Schroth

Bruno Streit

Bernd Schierwater*

Abteilung Ökologie & Evolution,

J. W. Goethe-Universität,

D-60054 Frankfurt,

Germany

**and Department of Biology,*

Yale University, New Haven,

Connecticut 06520-8114, USA

e-mail: Schierwater@zoology.uni-frankfurt.de

1. Carr, A. F. *BioScience* **14**, 49–52 (1964).
2. Lohmann, K. J. & Lohmann, C. M. F. *Nature* **380**, 59–61 (1996).
3. Bowen, B. W. & Avise, J. C. in *Conservation Genetics: Case Histories from Nature* (eds Avise, J. C. & Hamrick, J.) 190–237 (Chapman and Hall, New York, 1995).
4. Allard, M. W. et al. *Copeia* **1994**, 34–41 (1994).
5. Bowen, B. W. *BioScience* **45**, 528–534 (1995).
6. Williams, J. G. K. et al. *Nucleic Acids Res.* **18**, 6531–6535 (1990).
7. Hadrys, H., Ballick, M. & Schierwater, B. *Mol. Ecol.* **1**, 55–63 (1992).
8. Laurent, L. *WWF Report* (International Mediterranean Program RAC/SPA, Greenpeace Mediterranean Sea Project, 1994).
9. Bowen, B. W. et al. *Conserv. Biol.* **7**, 834–844 (1993).
10. Karl, S. A., Bowen, B. W. & Avise, J. C. *Genetics* **131**, 163–173 (1992).
11. Schierwater, B. & Ender, A. *Nucleic Acids Res.* **21**, 4647–4648 (1993).
12. Excoffier, L., Smouse, P. E. & Quattro, J. M. *Genetics* **131**, 479–491 (1992).

Erratum

In the Scientific Correspondence "Competition for royalty in bees" by R. F. A. Moritz, Per Kryger and Mike H. Allsopp (*Nature* **384**, 31; 1996), the y-axis labels of Fig. 2, showing the mean offspring frequencies of various patrines of a honeybee colony, were printed incorrectly. In the top panel the scale should have read '0–20%', not '0–2%' as printed, and in the middle and bottom panels '0–80%' rather than '0–8%'.

Human effect on global climate?

SIR — The recent pattern-correlation analysis of Santer *et al.*¹ has drawn considerable attention: Nicholls², for example, stated that it represents the "clearest evidence yet that humans may have affected global climate". This conclusion is based on the increasing similarity of the vertical temperature pattern in free-atmosphere and global-climate models incorporating combinations of carbon dioxide, sulphate aerosols and stratospheric ozone concentrations as measured by a 'centred' correlation statistic, $R(t)$.

Santer *et al.*¹ found significant increases in $R(t)$ in the 850–50-hPa region of the atmosphere over the period of their study (1963–87) for each of the climate model/atmospheric constituent combinations they analysed (with the exception of a sulphate-only model). We agree with Santer *et al.*¹ that this result stems largely from the pattern of stratospheric and upper tropospheric cooling and a hemispheric asymmetry in the lower and mid-troposphere. But we believe that the reported increases in $R(t)$ are almost totally caused by the higher-altitude temperature changes, and that in the lower levels the results are only a reflection of the time period chosen.

For example, using the carbon dioxide + sulphate model results from Taylor and Penner³ (Fig. 2 of ref. 1), we can compare the explained variance (obtained by squaring $R(t)$) at the end of the record (1987) between the 850–50-hPa layer (troposphere + stratosphere) to that for the 850–500-hPa layer (lower troposphere only). The values are 64% and 5%, respectively. Thus 92% of the explained variance results from the addition of data above 500 hPa. This result is typical for the models used in this study (with the exception of the sulphate-only model). Further, the calculation of $R(t)$, using four levels above 500 hPa and only two beneath, assures heavy dependence upon the upper troposphere and stratosphere.

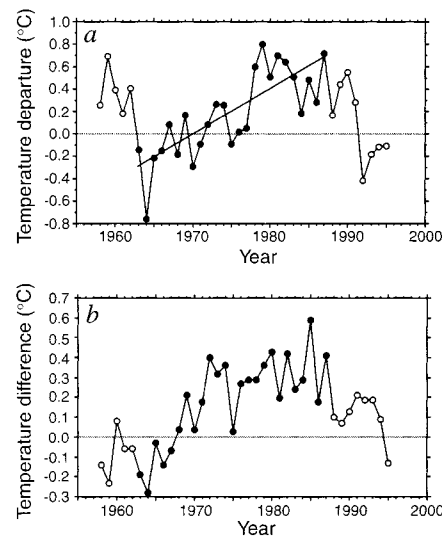
The factor that largely constitutes the hemispheric asymmetry is the warming of the mid-troposphere in the Southern Hemisphere, which appears both in the CO₂ + sulphate models (Fig. 1c, *e*, *g–i* of ref. 1) and in the observations, where it is concentrated between 850 and 300 hPa, 30 to 60° S (Fig. 1j of ref. 1). The observed data in ref. 1 are from the record of Oort and Liu⁴. This record extends to 1989 (although annual data to only 1987 are used by Santer *et al.*¹). The radiosonde record of Angell⁵ is longer (1958–95). Over the period of concurrency between refs 1 and 5, the correlation coefficient between annual anomaly values is 0.94 in the Northern Hemisphere and 0.92 in the Southern Hemisphere.

When we examine the period of record

used by Santer *et al.* in the context of the longer period available from ref. 5, we find that in the region with the most significant warming (30–60° S, 850–300 hPa) the increase is largely an artefact of the time period chosen (*a* in our figure). Although there is a statistically significant warming in the period from 1963–87, there is no significant change in the entire (1958–95) record. This has considerable bearing on the portion of $R(t)$ in ref. 1 that emanates from Southern Hemisphere mid-tropospheric warming. This result cannot be an artefact of the data, as precisely the same set of stations was used by Angell during the entire record.

Additionally, as $R(t)$ is a pattern-matching statistic, the increases in the tropospheric contribution to it are strongly dependent on the hemispheric temperature differences. The climate models predict that the Southern (sulphate-free) Hemisphere should warm relative to the Northern (sulphate-laden) Hemisphere, and for the period of record used in this study (1963–87) the observations agree. However, a close examination of upper-air temperature records again exposes this agreement as being fortuitous.

Using data from ref. 5, we present the 850–300-hPa hemispheric temperature differences (*b* in our figure). These data are highly correlated with the tropospheric $R(t)$ values calculated in ref. 1 from the CO₂ + sulphate model in ref. 3 ($r = 0.80$).



a, The Angell⁵ temperature departure (°C) history for 1958–95 for the 850–300-hPa layer between 30 and 60° S latitudes. Solid circles, years studied in ref. 1. There is a statistically significant ($P < 0.0001$) increase during 1963–87, whereas the overall record exhibits no significant trend. *b*, Annual temperature difference (°C) between Southern and Northern hemispheres in the 850–300-hPa layer, from ref. 5. There is a statistically significant ($P < 0.05$) downward trend in this data since 1972. Solid circles, period studied by Santer *et al.*

On the basis of this relationship, estimates of $R(t)$ for the period 1988–95, coupled with the observed values in ref. 1, yield a significant downward trend since the early 1970s. This is typical of the models tested.

Even though global temperatures were dramatically reduced by the eruption of Mount Pinatubo in the latter part of this period (1991), the hemispheric temperature difference was relatively unaffected. Therefore, we believe that the eruption would have little effect on the tropospheric values of $R(t)$.

In conclusion, we suggest that the increasing pattern correlation between climate models and observations found by Santer *et al.* is primarily governed by a signal in the upper troposphere and lower stratosphere, and that in the lower and mid-troposphere, the models and observations have been drifting further apart since the early 1970s. Such a result in the troposphere cannot be considered to be a 'fingerprint' of greenhouse-gas-induced climate change. It is therefore apparent that statements about the strength of the evidence for human alteration of lower tropospheric climate must be tempered in the light of more complete data than were analysed by Santer *et al.*

Patrick J. Michaels

Paul C. Knappenberger

Department of Environmental Sciences,
University of Virginia,
Charlottesville, Virginia 22903, USA

SIR — Santer *et al.*¹ present a statistical analysis in which they compare the model-generated zonal mean vertical thermal structure of the atmosphere in response to the concomitant increase of the concentration of greenhouse gases, sulphur emissions and the observed decrease in stratospheric ozone to the observed thermal structure of the atmosphere between 1963 and 1987. They find that the pattern correlation between the predicted and observed changes in zonal mean latitude height profiles of atmospheric temperature increases with time.

The authors attributed the largest amplitude signals of those trends primarily to two factors: first, the pattern comparison over 50 to 850 hPa (the signal of modelled tropospheric warming and stratospheric cooling); and second, in the troposphere, between 500 and 850 hPa, the disparity between Southern and Northern hemispheric warming due to modelled sulphate aerosol effects.

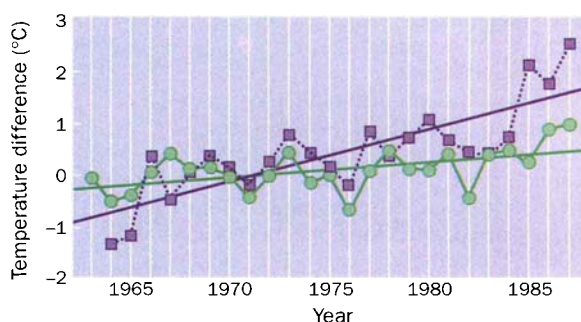
A well-known tropospheric and stratospheric temperature history for various latitudinal bands between 1958 and 1992 has already been published by Angell⁵; coverage for the Southern Hemisphere stratosphere begins in 1964. Angell's data of the layer 50–100 hPa closely correspond to the stratospheric data used by Santer *et al.*; similarly, Angell's 300–850-hPa data capture the bulk of the troposphere. A signal

of tropospheric warming and stratospheric cooling should be visible as an increasing difference between tropospheric warming and stratospheric cooling.

Taking the first point, a trend analysis of Angell's lower stratospheric data for the period 1963–87 (or 1964–87) shows a much larger cooling rate in the Southern than in the Northern Hemisphere (see figure). It appears, therefore, that most (stratospheric cooling in the Southern Hemisphere is about three times as large as in the Northern Hemisphere) of the signal pattern strength in 1963–87 relating to stratospheric/tropospheric trend differences originates in the Southern Hemisphere. This result seems to hold for the 1958–92 period as well, and for trends originating in the mid-1970s and later, which do not show any tropospheric warming in the mid-latitudes of the Southern Hemisphere, but sharply increasing stratospheric cooling and therefore increasing trend differences and strengthening of a pattern of stratospheric/tropospheric differences.

As Santer *et al.* point out, and as dynamical modelling results seem to suggest^{6–9}, the cooling of the Southern Hemisphere stratosphere, which increased sharply after about 1983 (ref. 5), may be related to stratospheric ozone depletion, which is most pronounced in the higher latitudes, but also occurs in the mid-latitudes of the Southern Hemisphere¹⁰. Therefore, the increasing signal pattern strength reported by Santer *et al.*¹ may primarily be related to Southern Hemisphere stratospheric cooling linked to ozone depletion due to CFCs (chlorofluorocarbons)^{6–9}. The possible human-induced climate effect alluded to by Santer *et al.* could then largely be attributed to stratospheric cooling by CFCs but not to the warming effect of anthropogenic greenhouse gases.

Turning to the second point, Santer *et al.* report a hemispheric-scale asymmetrical warming signal in 1963–87 only in those



Tropospheric and stratospheric temperature differences in the Northern (green symbols) and Southern (purple symbols) Hemispheres between 1963 and 1987, adapted from ref. 5. The least-squares linear trends (solid lines) are increases of 0.97 °C per decade in the Southern Hemisphere, and 0.27 °C per decade in the Northern Hemisphere.

modelling experiments that include cooling sulphate aerosol effects. The implication of this result is that the lack of sulphate emissions in the Southern Hemisphere has led to a larger tropospheric warming there.

This interhemispheric difference should be largest in the mid-latitudes, where most of the sulphur emissions occur (Fig. 1 of ref. 1). Although there is a larger 1963–87 warming in Angell's 300–850 hPa mid-latitude data (30–60 °C in both hemispheres) as well, it does not seem to be a permanent feature of the climate system in recent decades. A comparison between Northern and Southern Hemisphere trends ending in 1992 shows that trends beginning after 1964 imply increasingly greater warming in the Northern Hemisphere than in the Southern Hemisphere. This would be incompatible with increasing Northern Hemispheric cooling due to rising sulphur emissions there.

Regarding the role of natural factors, the early years of the period 1963–87 were substantially influenced by tropospheric cooling (and stratospheric warming) following the eruption of Mount Agung¹¹, whereas the end of that period was influenced by several strong El Niño events¹², which have led to some tropospheric warming and stratospheric cooling, particularly in the southern subtropics of the lower latitudes⁵. Therefore, the general tropospheric warming and stratospheric cooling trend between 1963 and 1987 has

- Santer, B. D. *et al.* *Nature* **382**, 39–46 (1996).
- Nicholls, N. *Nature* **382**, 27–28 (1996).
- Taylor, K. E. & Penner, J. E. *Nature* **369**, 734–737 (1994).
- Oort, A. H. & Liu, H. J. *J. Climatol.* **6**, 292–307 (1993).
- Angell, J. K. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A. *et al.*) 636–672 (ORNL/CDIAC, Oak Ridge, TN, 1994).
- Volk, C. M. *et al.* *Science* **272**, 1763–1768 (1966).
- Kiehl, J. T. *et al.* *Nature* **332**, 501–504 (1988).
- Atkinson, R. J. *et al.* *Nature* **340**, 290–294 (1989).
- Schwarzkopf, M. D. & Ramaswamy, V. *Geophys. Res. Lett.* **20**, 205–208 (1993).
- Stolarski, R. *et al.* *Science* **256**, 342–349 (1992).
- Angell, J. K. *J. Geophys. Res.* **93**, 3697–3704 (1988).
- Climate Diagnostic Bulletin August 1989* 6 (US Dept Commerce, Washington DC, 1989).
- Angell, J. K. *J. Climatol.* **1**, 1296–1313 (1988).
- Angell, J. K. *Geophys. Res. Lett.* **17**, 1093–1096 (1990).
- Gaffen, D. J. *J. Geophys. Res.* **99**, 3667–3676 (1994).
- Trenberth, K. E. & Olson, J. G. in *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations* (ed. Schlesinger, M. E.) 249–259 (Elsevier, Amsterdam, 1991).
- Parker, D. E. & Cox, D. I. *Int. J. Climatol.* **15**, 473–496 (1995).
- Christy, J. R. *Clim. Change* **31**, 455–474 (1995).
- Kalnay, E. *et al.* *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
- Tett, S. F. B., Mitchell, J. F. B., Parker, D. E. & Allen, M. R. *Science* **274**, 1170–1173 (1996).
- Wigley, T. M. L. *Nature* **339**, 365–367 (1989).
- Wigley, T. M. L., Jaumann, P. J., Santer, B. D. & Taylor, K. E. *Clim. Dynam.* (in the press).
- Ramaswamy, V., Schwarzkopf, M. D. & Randel, W. J. *Nature* **382**, 616–618 (1996).
- Wigley, T. M. L. & Raper, S. C. B. *Nature* **357**, 293–300 (1992).
- Kattenberg, A. *et al.* in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. *et al.*) 285–357 (Cambridge Univ. Press, Cambridge, 1996).

been accentuated by widely known natural factors^{13,14} and could at least partially be explained by them.

In conclusion, the main patterns described by Santer *et al.*¹ — interpreted by some as a sign of a human impact on the climate system — are either not permanent features of the climate system or cannot be ascribed to an increase of man-made greenhouse gases. The possible human impact on climate appears to be restricted to CFCs and the Southern Hemisphere stratosphere.

Gerd R. Weber

Gesamtverband des deutschen

Steinkohlenbaus,

Friedrichstrasse 1, 45128 Essen, Germany

e-mail: gvst_ure.weber@t-online.de

SANTER *ET AL.* REPLY — Michaels and Knappenberger, and Weber, in their contributions above, criticize our study¹ in which we attempted to identify human influences on climate.

Weber states that the increasing pattern similarity between a model signal and observed data (over 850–50 hPa) that we found may “largely be attributed to stratospheric cooling by CFCs...”. He bases this conclusion on larger stratospheric cooling in the Southern Hemisphere. Both Weber and Michaels and Knappenberger argue that the hemispheric asymmetry in 850–500-hPa temperature trends identified in our study is a transient feature unrelated to anthropogenic influences.

To support their arguments, Michaels and Knappenberger and Weber use the virtual temperature data set of Angell⁵, which has instrumental biases¹⁵ and known deficiencies in its spatial representativeness¹⁶. Angell’s large asymmetrical cooling in the lower stratosphere (greater cooling in the Southern Hemisphere) is not substantiated by analysis of other data sets — only a small long-term asymmetry trend is evident in the Parker radiosonde data¹⁷ (*a* in our figure). Furthermore, lower stratospheric temperature trends computed from satellite data¹⁸ and a reanalysis of operationally produced climate data¹⁹ show an asymmetry of the opposite sign (greater lower-stratospheric cooling in the Northern Hemisphere).

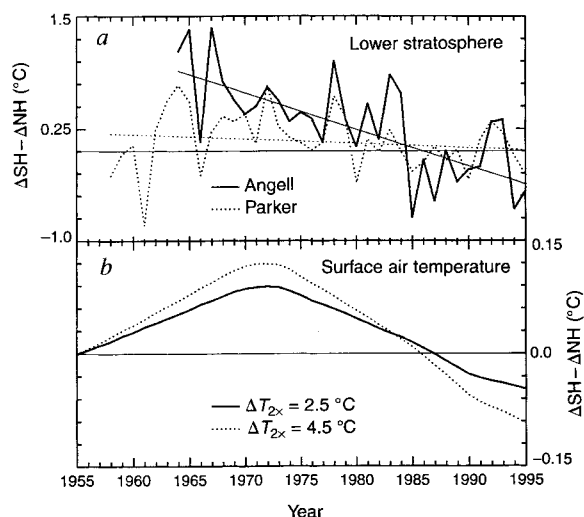
Thus, the basis for Weber’s argument for a dominant effect of CFCs is not supported by other available estimates of lower stratospheric temperature change. Furthermore, his proposed mechanism of stratospheric ozone depletion leads to a cooling of the troposphere²⁰, not a warming as observed. Our own¹ and more recent²⁰ work finds closest agreement between modelled and observed vertical temperature-change patterns when multiple anthropogenic forcings are considered.

Both Michaels and Knappenberger and Weber claim that our pattern-correlation (*R(t)*) results for the lower atmosphere are merely a manifestation of natural varia-

bility. To support this claim, Michaels and Knappenberger use the hemispheric temperature-change difference in the lower atmosphere (850–300 hPa) from Angell’s data. They contend that this time series is a reasonable ‘proxy’ for our 850–500 hPa *R(t)* results, and hence can be used to extend our correlation analysis beyond 1987.

To test this claim, we used the newly available Parker radiosonde data to extend our *R(t)* results for the low- to mid-troposphere to the period 1958–95. The time series of *R(t)* and hemispheric temperature contrast computed with the Parker data are highly correlated ($r=0.80$), thus confirming Michaels and Knappenberger’s supposition. Like the hemispheric temperature-change contrast in the Angell and Parker data sets, the ‘updated’ *R(t)* does decrease after 1988.

Contrary to Michaels and Knappenberger’s claim, however, such behaviour is consistent with our current understanding of anthropogenic causes. This is because there are temporal changes in the relative strengths of the greenhouse-gas and aerosol forcings, and in their associated (asymmetrical) climate response patterns. Thus, both *R(t)* and its ‘proxy’ are expected to show periods of increase and decrease (as in Michaels and Knappenberger’s figure *b*) as part of an anthropogenic signal^{21,22}. Model-based results for the hemispheric temperature-change contrast, based on anthropogenic forcing alone (our figure *b*), are qualitatively similar to the observations shown in Michaels and Knappenberger’s figure *b*. Thus, the decadal timescale fluctuations in both the hemispheric temperature differential and in *R(t)* most probably reflect an anthropogenic signal plus superimposed natural variability noise, and not noise alone, as Michaels and



Knappenberger, and Weber, contend.

In summary, the claim by Weber that our 850–50-hPa results reflect only CFC-related stratospheric ozone effects is incorrect and based on suspect data. Nevertheless, stratospheric ozone is an important component of the climate system, and its depletion may well contribute a significant part of the anthropogenic climate-change signal in the lower stratosphere²³. With regard to the claims in both contributions above that our results depend on the choice of data period, on the contrary, the use of a longer observed record fully supports our earlier 850–50-hPa results. For 850–500 hPa, the changes in *R(t)* are similar to changes in the hemispheric temperature contrast shown by Michaels and Knappenberger. This is not surprising: we ourselves interpreted our significant 850–500 hPa *R(t)* results primarily in terms of warming of the Southern Hemisphere relative to the Northern Hemisphere. However, both the recent change in hemispheric temperature contrast and the decline in *R(t)*, rather than being in conflict with the expected effects of anthropogenic forcing, are consistent with those expectations, and the primary conclusions of ref. 1 stand.

Benjamin D. Santer
and the 12 co-authors of ref. 1*
James S. Boyle

Program for Climate Model Diagnosis and Intercomparison,

Lawrence Livermore National Laboratory,
Livermore, California 94550, USA

email: bsanter@pcmdi.llnl.gov

David E. Parker

Hadley Centre for Climate Prediction and Research,

Meteorological Office,
Bracknell RG12 2SY, UK

*Affiliations as in ref. 1.

At ground zero

Daniel S. Greenberg

An Exhibit Denied: Lobbying the History of Enola Gay. By Martin Harwit. *Copernicus*: 1996. Pp. 477. \$27.50.

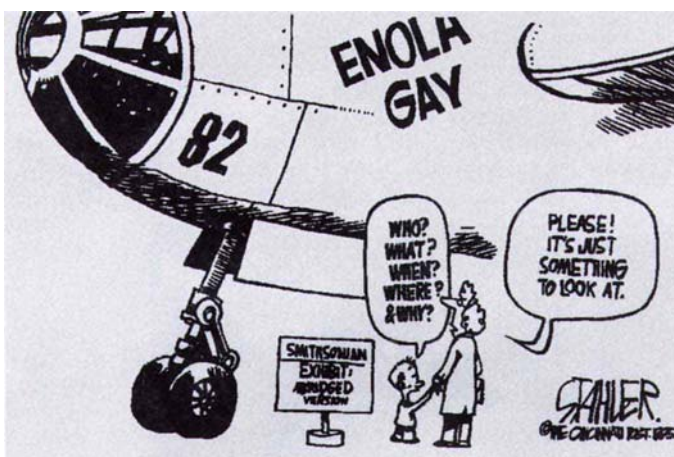
In 1987, Martin Harwit, professor of astronomy and historian of science and technology at Cornell University, was appointed director of the National Air and Space Museum, part of the Smithsonian conglomerate of museums in Washington DC. Eight years later, he was forced to resign, after a bitter collision between jingoist fervour and scholarly independence over the commemoration of the atomic bombing of Hiroshima.

In *An Exhibit Denied*, Harwit describes his experiences, memo by memo, meeting by meeting, mostly with an above-the-fray historian's dispassion, although occasionally with a dash of anger. For aficionados of the writhings of great bureaucracies under stress, internecine back-stabbing, reptilian politics and other pathologies of organizational behaviour, Harwit's account ranks as prize reading. The story he tells is of one of the darkest hours in the long history of the great Smithsonian.

On one side were Harwit and his museum colleagues, untutored in the harsh ways of strife and politics in the country's capital. On the other side were angered veterans of the Second World War, a strident press and influential politicians — all spouting hostility towards the leader and staff of the museum that prided itself on being the most popular in the world.

At issue were the museum's long-gestating plans for an exhibition to open in 1995 in observation of the fiftieth anniversary of the atomic bombing of Hiroshima, for which the centrepiece was to be the Enola Gay, the B-29 aircraft that dropped the bomb. A decade earlier, after years of neglectful storage, museum technicians had begun a painstaking restoration of the huge aircraft. The Enola Gay well satisfied the criterion of historic importance for inclusion in the Air and Space collection. But unlike virtually all the other hardware on display, the Enola Gay had not gained admission as a triumphant manifestation of higher, faster or farther. Rather, it had initiated the age of nuclear weapons, killed some 100,000 people and hastened the end of a horrible war. What were the museum's visitors to be told about this aircraft and its place in history?

Did the bombing avert a blood-bath invasion of Japan, as many surviving veterans gratefully believe to this day? Was it really a war-shortening, therefore life-saving, technological triumph, deserving of proud commemoration? Or, as asserted by a small group of historians, was the bombing an immoral, militarily superfluous



atrocities, inflicted on an already defeated Japan, possibly as a warning to a fractious Soviet Union?

These questions might normally be confined to the seminar table. But many Second World War veterans, including those who had served with the 509th Composite Group, the specially created unit that had dropped the bomb, were in their seventies, and were demanding an approving, if not celebratory, observation of the Enola Gay's historic wartime feat. As the leader of one of these increasingly impatient veterans groups put it, the plane should be displayed "proudly". Harwit did not disagree. But more than pride, he insisted, must be embodied in the exhibit.

For the decade of its existence before Harwit's appointment, the museum, in the words of Michal McMahon, a historian of technology, was "largely a giant advertisement for air and space technology". Harwit brought to his new job a dimension largely neglected under previous regimes — a willingness, perhaps an eagerness, to recognize the destructive side of aerospace technology, as manifested in its military applications. One of his first initiatives as director was to organize a study of strategic bombing under the title "From Guernica to Hiroshima: Bombing in World War II" — a linkage he quickly abandoned when veter-

ans and Air Force officials denounced the pairing of the Nazi bombing atrocity in Spain with the attack on Hiroshima. Recalling his harmonious discussions with an advisory committee in the early planning stages of the Enola Gay exhibit, Harwit writes: "To avoid the perception of cold, heartless militarism, we needed to infuse awareness that the bomb had caused damage and suffering".

Harwit went on to explain: "The Enola Gay was not just any B-29 bomber. She was the first bomber to have dropped an atomic bomb. That bomb had instantaneously killed close to 100,000 people. This was a fact that could not be suppressed without distorting history. On the other hand", he acknowledged, "this historic fact must also be placed in context."

And, on that basis, he proceeded with a doomed attempt to reconcile the diverse and uncompromising interests in the Enola Gay saga. Active and increasingly aggressive in their demands were veterans organizations, described by one member as "suspicious that the Smithsonian is more interested in sending a political message about strategic bombing and nuclear weapons". "Our position", he declared, "is that — like it or not — the

Enola Gay did make history." The veterans, influential with the Congress that finances the Smithsonian, were incensed by plans to display charred artefacts from the Hiroshima bombing, including a child's lunch box and a clock stopped at the moment of detonation. Critics said the planned exhibit sentimentally ignored Japanese culpability and cruelty in the war that began with a sneak attack on Pearl Harbor. Rattled by the mounting storm, one of the museum's senior scholars feared that the exhibit was "telling the events from the viewpoint of the victim".

Meanwhile, Japanese authorities were becoming queasy about lending artefacts of the Hiroshima and Nagasaki bombings to the Air and Space exhibit, and Harwit became concerned about igniting an emotional blow-up in Japan.

At times Harwit revealed extraordinary political naivety. Seeking financial support for the restoration of the Enola Gay, he approached the president of Nippon Television and raised the possibility of eventually displaying the historic B-29 in Japan. The offer was rejected with the explanation that "The Hiroshima and Nagasaki bombings remain firmly imprinted in the Japanese consciousness, much as the Holocaust does with the Jewish people".

As the struggle over historic interpretation intensified, Harwit found himself

under fire from Congress and lacking support from his Smithsonian superiors. He sought political cover behind a seeming endorsement of the initial exhibit plans by the director of the Center for Air Force History, Richard Hallion, who, while listing some reservations, described the plans as "a most impressive piece of work, comprehensive and dramatic, obviously based on a great deal of research, primary and secondary". Hallion added: "A bit of 'tweaking' along the lines discussed here should do the trick".

But each tweak by the increasingly isolated Harwit brought demands for greater concessions. And Hallion later denied that he had endorsed the exhibit script. At the demand of the American Legion, a reference to Japanese "defenders" in the battle of Okinawa was changed to "troops", because, Harwit explains, the Legion felt defenders "suggested sympathy for the Japanese".

In 1994, a new secretary, Michael Heyman, was appointed to head the Smithsonian, and the Republicans won

control of Congress. Along with other federal dependencies, the Smithsonian felt the budget lash. "With money the highest priority of the Institution, academic integrity began to take second place", Harwit writes.

In January 1995, Heyman cancelled the exhibit that was to surround the Enola Gay, leaving just the forward section of the fuselage — all that would fit into the museum — on display with scant reference to its historic role. Veterans and their families, he said, "were expecting, and rightly so, that the nation would honor and commemorate their valor and service. They were not looking for analysis, and, frankly, we did not give enough thought to the intense feelings such an analysis would evoke."

On request, Harwit promptly resigned a few months later. □

Daniel S. Greenberg is editor and publisher of *Science and Government Report*, 3736 Kanawha St NW, Washington DC 20015, USA.

Darwinism alive and well

James F. Crow

Selection: The Mechanism of Evolution. By Graham Bell. *Chapman and Hall*: 1997. Pp. 699. £55, \$75.

The Basics of Selection. By Graham Bell. *Chapman and Hall*: 1997. Pp. 378. £29.95, £37.50 (pbk).

Plan and Purpose in Nature. By George C. Williams. *Weidenfeld and Nicolson*: 1996. Pp. 191. £11.99. To be published in the United States by BasicBooks in February at \$20.

DARWIN started an avalanche; the books continue to pour forth. Do we need more? In the case of these three, yes. Graham Bell has written two volumes on selection, one comprehensive, the other an abridgement. The third book, by George Williams, is one of a series published under the general heading "Science Masters", intended to be responsible popularizations of science. It tells the story of adaptation in nontechnical language, easily accessible to the general reader.

Readers familiar with Bell's *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (University of California Press, 1982) will expect — and find — erudition, breadth, unusual diversity of examples and good writing in his latest works. The organization is innovative. In the manner of Martin Luther, Bell posts 175 numbered propositions in *Selection*. Some examples: 1, RNA viruses are the simplest self-replicators; 26, Characters other than fitness have intermediate optima; 126, Some genes distort meiotic segregation so as to favour their own transmission; 158, Resistance and virulence are costly; 162, Sexual selection and natural selection are antagonistic; 175, Sexual isolation may evolve as a correlated response to divergent selection in separate lines. In the main text each statement is followed by an elaboration, usually several paragraphs. Interspersed throughout are technical and experimental details, along with references to the literature. These are set off in boxes. Students should find this appealing. The boxes can be skipped on first reading and the numbered statements, conveniently listed at the beginning, provide an easy way to review.

Like R. A. Fisher's *The Genetical Theory of Natural Selection* (Dover, 2nd edn 1958), this book is explicitly about selection. In many ways it is the successor to the Fisherian bible. Yet it is very different: it is verbal, not mathematical; it has hardly any population genetics; it includes a great deal more experimental



Nor all the birds pictured in *Vanishing Songbirds* are endangered. The chipping sparrow (above) is still found in backyards and farmland across North America. This book is a portfolio of passerines in their natural habitats and represents the culmination of a lifelong study of American songbirds by renowned photographer Eliot Porter, who died in 1990. There are brief notes on each of the 125 species included. Bulfinch Press, \$37.50.

material; much of the emphasis is on microbes and agricultural plants; and the human species gets less attention.

The first experimental data in the book are from Sol Spiegelmann's clever experiments with virus Q β , in which selection for rapid replication reduced the organism to its bare minimum. Of course, multicellular and diploid species and the complications of Mendelism are brought in, but later. Kin and group selection, evolution of senescence, life cycles, transposable elements, sexual selection and speciation are thoughtfully and clearly discussed. Many readers will be surprised by the insights from chemostats, from *Chlamydomonas* (much is Bell's own work) and from agricultural plants.

This is not a book about evolution, but solely and unabashedly about selection. Phylogeny and molecular evolution get short shrift. Traditional population genetics, with the usual formulae and emphasis on diploid, sexual species, is largely ignored. Kimura's neutral theory is mentioned, only to be dismissed as irrelevant to the purpose of the book. Mutation, migration, random drift and population structure are considered only as they provide the background for or otherwise affect selection. Bell makes the limitation clear from the start: "This book has been written to make a point and to fulfil a need. The point is that the importance and the distinctiveness of the process of selection have been undervalued by most biologists. There is, consequently, the need for a book that describes the principles of selection in a simple but reasonably comprehensive way." I am not sure that selection has been so badly undervalued, but in any case a big book devoted to selection is most welcome.

Bell does not hesitate to express opinions, but he is fair. His own "tangled bank" theory he now regards as quantitatively inadequate. After stating his view that mating-type genes arose from transposable elements, he adds: "This view is not widely accepted".

I thought of some studies that might have been included, but it would be churlish to emphasize these when there are far more in the book that were new to me. What makes this book useful is its singleness of purpose together with the great breadth of coverage of the chosen subject.

Bell's second book reviewed here is an abridgement of the first, intended for the reader with less technical background or interest, or a shorter attention span. Some of the discussion is eliminated, as are all the boxes with references and technical details. The 175 propositions are all there, however, so it is easy for the reader to move to the full book for amplification and further edification.



THE shells pictured in *Shells: Treasures of the Sea* by Leonard Hill are not necessarily found on the seashore. These Hawaiian tree snails are threatened by habitat destruction and predation by the introduced snail *Euglandina rosea*. The book's 300 photos show shells from around the world in all their striking diversity, with information on biology, and how human cultures have used them in decoration. Hugh Lauter Levin Associates (distributed in the United Kingdom by Gazelle Book Services), \$75, £45.

Whereas Bell gives little attention to adaptation, except insofar as it fits into the discussion of natural selection, Williams makes it the central theme. An alternative title, he says, might have been "The Adaptationist Program". Like Paley, he admires the intricacy and effectiveness of countless structures and functions. But he goes further and emphasizes the constraints on evolution imposed by physics and chemistry and by past evolution. We should not expect to find intelligent engineering design. The human eye is one of his examples. It works very well, of course, but having nerves between the light source and the retina with the consequent blind spot could have been prevented by following the squid's lead. Another example is the human vas deferens, contorted into a hairpin loop by history and embryology. And surely any humane designer would have placed the prostate gland somewhere else.

Williams is widely known for his 1966 book *Adaptation and Natural Selection* (Princeton University Press, 1966), which put an end to much naive group-selection thinking. In this book, adaptation and the means by which it is brought about are treated clearly and in an elementary way. Williams is fond of analogies and uses them to good effect. There are lucid discussions on the evolutionary advantages of sexual reproduction, on senescence, on group selection and on philosophical implications. An important tenet is that medical education, research and practice could profit by guidance from evolutionary principles. Williams is concerned with the mismatch between Stone Age adaptations and the demands of modern society. The better we understand this, he

says, the better we can deal with the difficulties entailed.

Williams gives his own opinions freely, but he is very generous to those whose work he admires, such notables as G. G. Simpson, William Hamilton, Richard Dawkins, Ernst Mayr and John Maynard Smith. E. O. Wilson's *Sociobiology* is described as "probably the most philosophically important biological work of this century". By contrast, he is critical of Stephen Jay Gould. (Gould has written so much that it is hard to be critical of *all* that he says.) In contrast to Gould's disparagement of the adaptationist programme, Williams uses it as a guiding principle. "For each attribute of an organism, the program's practitioners raise the question: How does this relate to the organism's efforts to survive and pass on its genes?"

Who should read these books? Anyone with even a casual interest in evolution can enjoy and profit by Williams's book. It can be read like a novel, a novel of ideas. It is a great way to find out what a leading evolutionist is thinking about. Bell's books have a different role. Students who want to strip evolution of all else and concentrate on its special moulding force, selection, can do no better than start with Bell's short book. Those who want more can consult his big book, and specialists will find this a fine reference source. They might also get some good ideas for research. Even a card-carrying population geneticist like me found it very satisfying. □

James F. Crow is at the Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA.

Oil in troubled waters

Peter Kassler

Perspectives of Oil and Gas: The Road to Interdependence. By Marcello Colitti and Claudio Simeoni. *Kluwer*: 1996. Pp. 164. \$71.50, £48.

In 1866, the British statesman William Gladstone asked: "How long will the supply of energy last? What will be its effect on our economic life?" He was talking about coal. Today we are mainly concerned about oil and gas, but the same questions are no less relevant.

Those who would like to know more about the inner workings of the oil and gas industries should find this a useful book. The authors, respected Italian industry figures, begin with a brief but clear account of

be working much more seriously on the next wave of energy supplies, probably renewables or nuclear, or both. It is disappointing that the book gives little sense of the profound and growing influence of the environmental movement on oil and gas.

On the political front, the authors lead us deftly out of the smooth waters of the Seven-Sisters-managed 1960s oil world into the maelstrom of the 1970s and early 1980s with its massive economic upsets and transfers of power and resources.

With hindsight, well-meaning but incompetent interventions by governments of oil-producing and major consuming countries can now be seen to have transformed an important traded commodity into a political football. When kicked wildly about, this inflicted severe bruises on all the players, particularly on poor developing countries without energy resources. We now have a much calmer oil market, to a large degree de-politicized.

That said, the authors show a remark-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Crises in the pipeline: the oil market has suffered many political shocks.

the location, size and development of the world's oil and gas resources. This is followed by an equally concise description of the successive alarms, shocks and restructurings that have affected these industries, and therefore the global economy, in the past quarter century.

The oil story is told with an occasional Italian spin. We are reminded that, after ENI was excluded from its concessions by the Anglo-Saxon multinationals (the "Seven Sisters") in the 1960s, it was Enrico Mattei who took revenge by accepting 50:50 government profit-sharing in the Middle East. He may well have been the butterfly (in Milan on this occasion) whose wing-flapping triggered the first oil shock.

The authors' answer to Gladstone's question is reasonably reassuring. Oil and gas are sufficiently abundant to permit economic growth at today's rates until at least 2030. But the size and importance of these industries are such that we should already

able nostalgia for a return to government orchestration of oil supplies in order to do away with the 'casino' aspects of oil trading. I, for one, would frankly prefer to rely on a vigorous free market to keep my children's toes warm in their old age.

The Middle East remains at the centre of this story, and the authors correctly draw attention to the importance of its resources. In another recent book, *The New Geopolitics of Energy* (Royal Institute of International Affairs, London, 1996), John V. Mitchell reminds us that we will have to continue to live with that region's occasional political and military disruptions. It gives a less technical but more politically analytical account of post-Cold War oil and gas, and is recommended as a parallel volume to Colitti and Simeoni's valuable book. □

Peter Kassler, a consultant on energy and the environment, is at Manorwood, Cranleigh Road, Womersley, Surrey GU5 0PB, UK.

Christmas goodies

The Edge of the Unknown: 101 Things You Don't Know About Science and No One Else Does Either by James Trefil.

Houghton Mifflin, \$23. Great scientific questions such as "How did life begin?" or "So where are all the black holes?" or "Who were the earliest humans?" tackled by one of the great scientific explainers. Each essay is of no more than three pages and is self-contained. Authoritative, witty and well-organized, this compact volume is an ideal bedside book for the curious lay person.

The Cambridge Illustrated History of Medicine edited by Roy Porter.

Cambridge University Press, £24.95, \$39.95. A team of experts looks at the history of disease, health and medicine over the past 2,000 years or so, and puts into historical context modern developments and anxieties.

Congo Journey by Raymond O'Hanlon. Hamish Hamilton, £18. A past natural-history editor of *The Times Literary Supplement* tells of his epic cryptobiological journey into the depths of the Congo jungle in search of a dinosaur in a lost prehistoric lake.

Flora Britannica by Richard Mabey. Sinclair-Stevenson, £30. Covers the native and naturalized plants of England, Scotland and Wales. Richly illustrated with more than 500 photographs, this is more a cultural than a botanical flora: Mabey, one of Britain's leading writers on natural history, spent five years canvassing ordinary people about their memories, anecdotes, observations and regional knowledge. The result is a fascinating compendium of the roles of wild plants in social life, arts, custom and landscape.

First Light: The Search for the Edge of the Universe by Richard Preston.

Random House, \$24. A new edition of one of the modern classics of popular science writing: a thrilling and absorbing account of the astronomers at the Palomar Observatory, California, and their work there with the Hale telescope.

Our Scientific Heritage: An A-Z of Great Britain and Ireland by Trevor I.

Williams. Sutton, £20. A companion guide to the geographical links with British and Irish achievements in science, technology, archaeology, medicine and engineering. The author, a distinguished historian of science who died in October, covers some 750 sites, ranging from Brunel's suspension bridge in Bristol to Newton's home at Woolsthorpe. Also included are maps, a bibliography and an index of people, highlighting places of interest for each.

Martin Bond/SPL

Host factors and the pathogenesis of HIV-induced disease

Anthony S. Fauci

The level of human immunodeficiency virus (HIV) replication in patients reflects a balance between stimulatory and inhibitory host factors (particularly endogenous cytokines). New information concerning the cellular co-receptors for HIV and the cellular tropism of different strains of virus will advance our understanding of HIV-induced pathogenesis and suggests new therapeutic and preventive strategies.

THE pathogenesis of HIV-induced disease is complex and multifactorial^{1,2}. Following primary infection with HIV, a burst of virus replication disseminates the virus to lymphoid organs¹. A robust cellular and humoral immune response usually inhibits viral replication within weeks³⁻⁶ but the virus almost invariably escapes from immune containment, producing a chronic, persistent infection, leading to advanced clinical disease with a strikingly high mortality over a median period of about 10 years (ref. 7). Some viral replication occurs in lymphoid tissue (and perhaps elsewhere, for example in the brain) throughout the entire course of infection, even when the patient is clinically asymptomatic⁸. Indeed, sensitive techniques for detecting viral RNA reveal virus in the plasma at every stage of disease⁹. Although at any point in the infection, plasma viral levels are apparently stable, the virus is in fact turning over very rapidly^{10,11}, and $\sim 10^{10}$ virions are produced and cleared from the circulation each day^{12,13}. More than 99% of virus is thought to be produced by newly infected CD4⁺ T cells, whose half-life is 1.6 days (refs 10-12).

The steady-state level of plasma viraemia from about 6 months after infection strongly predicts how fast the disease will progress¹⁴ and has considerable pathophysiological significance. It is influenced by both host and viral factors. Among the latter are replicative fitness, predilection for mutation, and cytopathicity^{13,15}, all of which are important components of any model of HIV pathogenesis. This review, however, will focus on the role of host factors, particularly cellular activation and the role of cytokines and their receptors in regulating viral replication and in pathogenesis. Clearly, the immune response to HIV also affects the progress of the infection. This issue has recently been reviewed¹⁶ and will not be addressed here; instead, I will focus on the way in which the balance between HIV-inducing and HIV-suppressing host factors controls the net level of viral replication. Perturbation of this balance both *in vitro* and *in vivo* strongly affects the net replication of HIV and can significantly influence the course of disease (Fig. 1).

Exogenous factors and HIV replication

HIV replicates more efficiently in activated cells¹⁷⁻²⁰, and viral levels consistently increase when the immune system of HIV-infected individuals is activated by exogenous stimuli such as opportunistic pathogens²¹⁻²⁵. Active infection by *Mycobacterium tuberculosis*, for instance, substantially increases plasma viraemia, which returns to baseline when the infection is successfully treated²². This increase in the rate of viral replication is associated with cellular activation and expression of HIV-inducing cytokines, as well as an acceleration in the course of HIV-induced disease^{26,27} (see below). Similarly, the course of HIV infection appears to be more aggressive in sub-Saharan Africa than in developed countries²⁸⁻³¹, probably reflecting the immune activation associated with frequent and chronic infection by parasites and other pathogens^{32,33}. This immune activation is associated with

elevated levels of HIV-inducing cytokines such as TNF- α and interleukin (IL)-1 β and IL-6 (refs 33, 34).

The impact of immune activation on viral replication and disease progression has been confirmed and expanded in experimental studies of simian immunodeficiency virus (SIV)-infected monkeys and HIV-infected chimpanzees and humans³⁵⁻³⁷. HIV-infected individuals manifest transient bursts of viraemia on immunization with antigens such as influenza and tetanus toxoid³⁵⁻³⁸, and peripheral blood mononuclear cells (PBMCs) obtained from such individuals during the transient period of cellular activation following immunization support greater HIV replication *in vitro*³⁷. Immunization with tetanus toxoid also renders PMBCs from uninfected individuals more susceptible to infection with HIV *in vitro*³⁷. Exogenous factors thus play a major role in regulating viral replication, predominantly by perturbing cellular activation and cytokine production.

Endogenous cytokines

The immune system is regulated by a complex network of pleiotropic and redundant cytokines, which are continually secreted to a greater or lesser degree even when the system is apparently quiescent, and are further expressed to varying degrees in response to antigen³⁹. As HIV directly infects cells of the immune system and triggers a robust immune response, it is an important and persistent source of immune activation¹. This activation is intimately linked to cytokine secretion³⁹, and expression of several cytokines is enhanced in HIV-infected individuals. Secretion of the proinflammatory cytokines TNF- α , IL-6 and IL-1 β is increased in PBMCs⁴⁰, and expression of these cytokines together with interferon (IFN)- γ is greatly increased in lymphoid tissue^{40,41}. Elevated levels of TNF- α and IL-6 are also found in the plasma and cerebrospinal fluid⁴⁰.

Cytokine patterns have been linked to a polarization of the cell-mediated and humoral immune responses⁴². T-helper (Th)-1 cells are characterized by secretion of IL-2 and IFN- γ , and favour cell-mediated immune responses; Th-2 cells are characterized by secretion of IL-4, IL-5 and IL-10, and favour humoral immune responses^{42,43}. Although an imbalance in subsets of CD4⁺ T lymphocytes has been reported in HIV-infected individuals, with disease progression reflecting a transition from Th-1 to Th-2 cytokine patterns^{44,45}, this has been disputed^{41,46}; in humans several other cell types secrete these cytokines⁴¹, further confusing the issue. The subject has been discussed in detail elsewhere^{42,43}, and is beyond the scope of this review.

The finding that crude supernatants from PBMC cultures induce viral expression in chronically infected cell lines⁴⁷ led to the discovery that numerous individual cytokines induce HIV expression either when manipulated endogenously or when added to acutely or chronically infected cell cultures⁴⁰. Such cytokines include IL-1 β , IL-2, IL-3, IL-6, IL-12, TNF- α and TNF- β , and colony-stimulating factors M-CSF and GM-CSF. IFN- α and

IFN- β suppress HIV replication, whereas TGF- β , IL-4, IL-10, IL-13 and IFN- γ either induce or suppress HIV expression, depending on the culture system⁴⁰. Several of the active cytokines are synergistic, whereas others exert their effects in an autocrine and paracrine manner in the same way that they control the immune system⁴⁰. Both the *in vitro* study of PBMCs acutely infected with HIV and studies of PBMCs and lymph-node mononuclear cells from HIV-infected individuals indicate that HIV replication is tightly controlled by these cytokines. The blockade of cytokines such as TNF- α , IL-1 β and IFN- γ by receptor antagonists, or by antibodies against the factors or their receptors, consistently and occasionally completely suppresses viral replication^{48,49}.

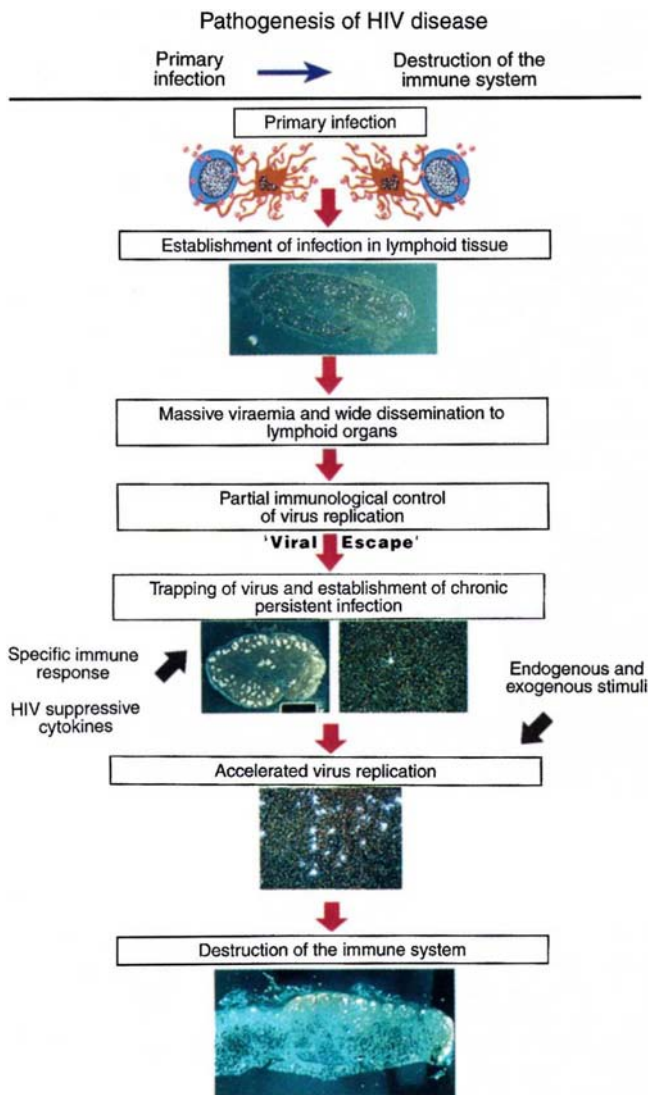


FIG. 1 Pathogenic events in HIV disease. On entry into the body, HIV probably encounters a dendritic cell, which is either infected or carries the virus to a CD4⁺ T cell in a lymphoid organ. Infection is then established in the lymphoid organ and a burst of viraemia seeds virus throughout the body. Viral replication is partially controlled, but the remaining replication almost invariably leads to escape from immune control, and establishment of chronic persistent infection in lymphoid tissue. The net level of viral replication is regulated by a balance between factors that induce virus replication and factors that contain or suppress it. In most patients, replication accelerates over a period of time, leading to the destruction of the immune system. The white dots in the lymphoid tissue represent individual cells expressing HIV viral RNA, as determined by *in situ* hybridization. Trapped virus (large clumps of white staining in the peripheral regions of the lymph node) is extracellular.

Another example of the effects of endogenous cytokines on HIV replication is that of IL-10, an immunosuppressive cytokine that inhibits HIV replication in acutely infected monocyte/macrophages by blocking secretion of the HIV-inducing cytokines TNF- α and IL-6 (ref. 50). *In vivo* infusion of a single bolus of IL-10 into normal individuals renders their cells unable to secrete cytokines in response to lipopolysaccharide and makes them relatively resistant to *in vitro* infection with HIV (A.S.F., D. Weissman, unpublished observations). In preliminary studies of HIV-infected individuals, a similar bolus dramatically decreased plasma viraemia for several hours, during which time their cells could not be induced to secrete TNF- α or IL-1 β *in vitro* (A.S.F. and D. Weissman, unpublished observations).

The molecular mechanisms by which these cytokines affect HIV replication are best understood for TNF- α , arguably the most important and certainly the most potent of the HIV-inducing cytokines⁴⁰. TNF- α activates the cellular transcription factor NF- κ B, which induces expression of numerous cellular genes⁵¹. Binding sites for NF- κ B occur within the long terminal repeat (LTR) of the HIV genome, and the effects of TNF- α on HIV are mediated by NF- κ B-induced transcription^{52,53}. Although IL-1 β is also thought to affect viral transcription, it is unclear whether NF- κ B is involved⁴⁰. Other cytokines, such as IL-6, GM-CSF and IFN- γ , have predominantly post-transcriptional effects⁴⁰.

Suppression of HIV replication

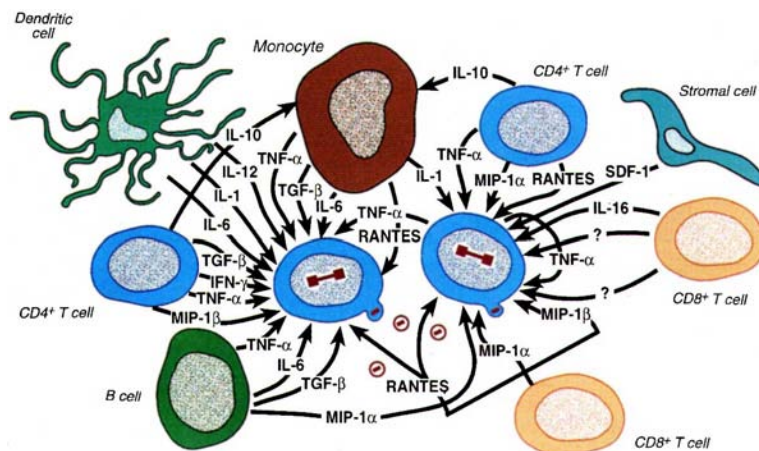
The effects of HIV-stimulating cytokines are counterbalanced by those that suppress HIV replication. An initial report of non-cytolytic suppression of HIV replication by CD8⁺ T cells^{54,55} was subsequently shown to reflect the action of a soluble factor whose precise identity remains unclear⁵⁵. The phenomenon has since been corroborated by several laboratories⁵⁵, however, and multiple suppressor factors whose activities can be distinguished in various *in vitro* models of HIV replication have now been identified^{56–58}.

The search for the identity of these factors has now triggered a new area of HIV research. In late 1995, Cocchi *et al.* showed that the suppressive effects of supernatants from CD8⁺ T cells transfected with human T-lymphotropic virus (HTLV)-1 or from HIV-infected individuals were due to the chemoattractant cytokines (β -chemokines) RANTES, MIP-1 α and MIP-1 β (ref. 58). The suppressive effects were observed in acute infections with some but not all strains of HIV-1, a dichotomy that was then unexplained. Although the β -chemokines were clearly the predominant suppressive factors secreted by CD8⁺ T cells in these experiments, they could not account for all of the suppressive effects mediated by CD8⁺ T-cell supernatants in other systems. In cultured macrophages, the β -chemokines did not suppress acute infection^{59,60}, whereas supernatants from CD8⁺ T cells did⁶⁰. Moreover, CD8⁺ T cells are not the only mononuclear cells that secrete RANTES, MIP- α and MIP-1 β (ref. 61). Indeed, HIV replication in cultures of CD8-depleted PMBCs from HIV-infected individuals has been shown to reflect the balance of endogenous HIV-inducing and HIV-suppressing cytokines⁴⁹. When the most potent HIV-inducing cytokines, such as TNF- α and IL-1 β are neutralized, viral replication is markedly decreased, and when the β -chemokines RANTES, MIP-1 α and MIP-1 β are neutralized in the same cultures, viral replication is increased. Addition of autologous CD8⁺ T cells to these cultures suppresses viral replication, but this suppression can only be overcome by neutralizing the β -chemokines when very low numbers of CD8⁺ T cells are added⁴⁹, indicating that the CD8⁺ T cells produce further unidentified HIV-suppressor factors. The steady state of HIV replication thus reflects the actions of various endogenous cytokines derived from multiple cell sources and can be affected by modulating these cytokines (Fig. 2).

Viral co-receptors and cellular tropisms

The ability of HIV-1 to infect different types of cells varies from

FIG. 2 Endogenous cytokines regulating viral replication in CD4⁺ T cells. Numerous cytokines, particularly the proinflammatory cytokines TNF- α , IL-1 β and IL-6, strongly upregulate replication. TGF- β and IL-10 downregulate it—in IL-10's case, at least in part, by downregulating proinflammatory cytokines. The β -chemokines, which are secreted by a variety of cell types including CD8⁺ and CD8⁻ mononuclear cells, strongly inhibit infection by M-tropic strains of HIV-1, whereas SDF-1 inhibits infection with T-tropic strains.



isolate to isolate and is referred to as cellular tropism. All strains of HIV infect primary CD4⁺ T lymphocytes. Many primary isolates also replicate well in monocytes, but not in transformed T-cell lines⁶², and are classified as monocyte-or macrophage-tropic (M-tropic), although they can also infect primary T cells. Other isolates, particularly those that have been passaged in lymphoid cells *in vitro* (T-cell-line-adapted strains), infect primary T lymphocytes but not monocytes or macrophages, and are referred to as T-cell-tropic (T-tropic) viruses. As HIV infection progresses, the initially dominant M-tropic viruses are usually replaced by T-tropic viruses^{63–65}. The former are generally found not to induce syncytia formation *in vitro*, although the latter do^{66,67}, and the terms M-tropic and non-syncytia-inducing (NSI) or T-tropic and syncytia-inducing (SI) are often used synonymously. This may not be entirely correct, as the transition from M-tropic/NSI to T-tropic/SI viruses as disease progresses may produce dual-tropic viruses that are M-tropic, despite SI characteristics^{68–70}. The viral determinant of cellular tropism maps to the gp120 subunit of the HIV-1 Env protein, particularly the third variable region or V3 loop of gp120 (ref. 71, see below). Studies to delineate the basis of cellular tropism led to the identification of the first bona fide co-receptor for HIV-1.

In early 1996, Feng *et al.* reported that the receptor CXCR4 (also termed fusin or LESTR) was the co-receptor responsible for the efficient entry of T-tropic strains of HIV-1 into target cells⁷². CXCR4 was then an 'orphan' receptor, in that its natural ligand was unknown; its sequence indicated that it was a G-protein-coupled receptor with seven transmembrane helices, most closely related to the receptor for the CXC chemokine IL-8. Curiously, antibodies against CXCR4 blocked the entry of T-tropic but not M-tropic strains, suggesting that their effects were the inverse of those of the β -chemokines, which block infection with M-tropic but not T-tropic strains³⁸ (see above). Taken together, these observations suggested that the β -chemokines inhibited infection with M-tropic strains by downregulating or blocking a second co-receptor. The β -chemokine receptor CCR5 can bind all three inhibitory chemokines⁷³, and a cluster of reports soon showed that it was indeed the co-receptor for M-tropic HIV-1 (refs 59, 70, 74–76). Other β -chemokine receptors (CCR2b and CCR3) can also act as co-receptors for M- and dual-tropic HIV, but they are little used by M-tropic strains⁷⁰. Shortly afterwards, the CXC chemokine SDF-1 was shown to be the ligand for CXCR4 and to block infection with T-tropic but not M- or dual-tropic HIV-1 (refs 77, 78).

Different strains of HIV-1 thus use different chemokine receptors as co-receptors, and infection by any one strain is blocked by the ligand for the receptor in question (Fig. 3). Unlike cytokines such as TNF- α , IL-1 β and IL-6, which enhance HIV replication by inducing viral gene expression⁴⁰ (see above), the β -chemokines seem to inhibit infection by blocking or downregulating the receptor they share with the virus⁷⁹ (K. Pettrone, D. Weissman

and A.S.F., unpublished observations), although direct and indirect effects on signal transduction may also play a part. Further studies are thus required, particularly as the complexity of the chemokine-receptor family⁷³ suggests that other receptors (some perhaps not yet unidentified) may play a role in HIV entry.

Co-receptors and resistance to infection

Some individuals repeatedly exposed to HIV infection nevertheless remain uninfected, including certain commercial sex workers⁸⁰ and sexual partners of HIV-infected individuals⁸¹. In studies of such individuals whose PBMCs were variably resistant to infection with certain strains of HIV *in vitro*⁸², cells from two individuals in particular were found to be highly resistant to infection with M-tropic virus, but readily infected with T-tropic virus. The gene for the major M-tropic receptor CCR5 was therefore sequenced, and both individuals were found to be homozygous for virtually identical 32-base-pair deletions⁸³, producing a truncated version of the receptor that was not expressed at the cell surface. PBMCs from these individuals did not transduce CCR5-mediated signals, although responses to other chemokines were normal. Both individuals appeared to be phenotypically fit with no host defence defects, suggesting that therapeutic strategies aimed at downregulating the receptor and/or blocking its ability to bind virus would not be deleterious (see below). Subsequent studies of HIV-1-infected cohorts have revealed no homozygotes among them^{83,86}.

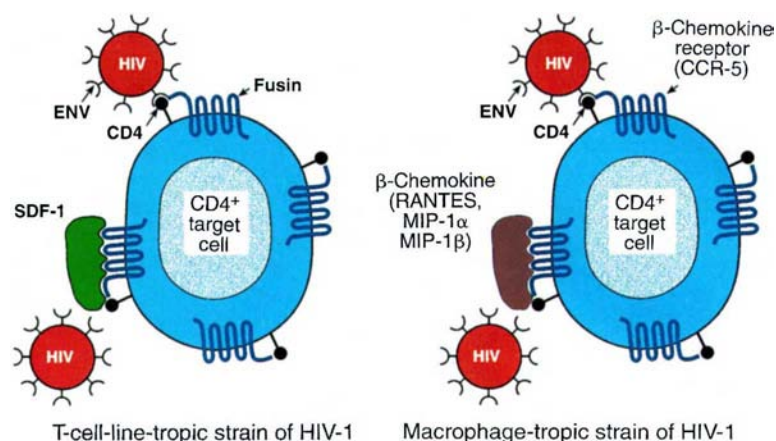
About 20 per cent of western European caucasians are heterozygous for the defect^{83,84}, but it has not been found in 124 western and central Africans or in 248 Japanese, indicating that it is either absent or extremely rare in these populations⁸⁴. The frequency of heterozygosity in a cohort of HIV-1-infected caucasians was 35 per cent less than in the general population, suggesting that heterozygosity may confer partial protection against infection⁸⁴, although this was not confirmed in another study⁸⁶. Moreover, PBMCs from the parents of homozygous individuals replicated virus less efficiently than PBMCs from normal individuals *in vitro*⁸³. Heterozygosity thus seems to provide a variable degree of protection against disease^{85,86}.

These findings suggest a simple model, in which the different receptors and their ligands explain the transmission and pathogenesis of T- and M-tropic strains of HIV-1, which would form an attractive basis for therapeutic and preventive strategies. But years of experience in the study of HIV suggest that even the simplest models of HIV pathogenesis can give rise to extraordinary complexities, and such models will almost certainly appear somewhat simplistic and require modification as new information becomes available.

Implications

Cellular tropism. The different receptors used by T- and M-tropic strains of HIV-1 have major implications for the basis of cellular

FIG. 3 Co-receptors for M- and T-tropic strains of HIV-1 on CD4⁺ T cells. CXCR4 (fusin, LESTR) is the major co-receptor for T-tropic strains, entry of which is inhibited by its ligand SDF-1. CCR5 is the major co-receptor for M-tropic strains, and their entry is inhibited by its ligands RANTES, MIP-1 α and MIP-1 β . Primary T cells and monocytes express both co-receptors; the former are susceptible to both strains, but the latter can only be infected by M-tropic strains, for reasons that are not yet clear.



tropism and the mechanisms by which the virus enters target cells. Viral entry requires binding to a target cell, followed by a series of receptor-triggered conformational changes in the major viral envelope protein (Env)^{59,70–72,74–76,87,88}. CD4 alone can induce conformational changes in Env, but these are insufficient for entry⁸⁹. Several roles have thus been proposed for the newly described co-receptors, including the induction of further changes in Env by a direct or indirect association with the co-receptor, or direct binding of Env to the co-receptor, bringing the viral protein to the cell membrane to allow fusion⁸⁸. Indeed, binding of HIV to CD4 greatly increases the efficiency of the gp120–CCR5 interaction^{90,91}, and there is evidence for physical association between CXCR4 and the CD4–gp120 complex in the cell membrane⁹². Chemokine-induced signalling may also depolarize or destabilize the cell membrane, assisting fusion⁸⁸, but neither G_i-protein signalling nor tyrosine kinase activity is required to inhibit M-tropic virus infection by RANTES (refs 79, 93 and K. Pettrone, D. Weissman and A.S.F., unpublished observations). Additional studies are required to delineate these events.

Viral transmission and host resistance to infection. These findings also have profound implications for the mechanisms of viral transmission and host resistance⁹⁴. As already mentioned, M-tropic strains of HIV-1 predominate shortly after infection^{64,65}, and these strains are responsible for initial infection⁶⁵ even when the transmitting partner harbours both M- and T-tropic viruses⁹⁵. The apparent resistance to infection conferred by homozygous defects in the CCR5 gene^{83–86} argues strongly that initial entry of M-tropic strains into susceptible target cells is critically important for establishing HIV infection. Why this should be so is still unclear. Primary infection may be confined to cells expressing CCR5 rather than CXCR4 or other potential co-receptors, selecting M-tropic strains from among the different strains deposited at the site of exposure. Such cells may predominate at the mucosal surfaces through which the virus is sexually transmitted. Moreover, the cytokine microenvironment of the tissue in question and its state of activation may modulate susceptibility to infection by affecting receptor expression. IL-2 upregulates CCR1, CCR2 and CCR3 (ref. 96), and other immunoregulatory or proinflammatory cytokines may affect expression of both CCR5 and CXCR4 (ref. 97). The natural co-receptor ligands may also block and/or downregulate their respective receptors, as well as affecting the expression of any other co-receptors involved in HIV-1 entry. In addition, inflammation at the site of viral exposure may tip the balance between expression of the co-receptor and the corresponding ligand in favour of viral entry, explaining (at least in part) the more efficient transmission to individuals with a concomitant sexually transmitted disease producing genital inflammation²⁸.

Interestingly, M-tropic viruses also predominate early in the infection of individuals directly inoculated with virus (such as injecting drug users and recipients of contaminated blood pro-

ducts), suggesting that the aforementioned selection process may not be restricted to mucosal surfaces^{98,99}. Initial infection in the reticuloendothelial system and lymphoid organs may also require cells such as monocyte/macrophages that are largely resistant to T-tropic viruses. After intravaginal inoculation of rhesus macaques with SIV, Langerhans cells beneath the mucosa are the first cells to be infected and probably initiate a spreading infection by migrating to the draining lymphoid tissue¹⁰⁰. It will be interesting to know which co-receptors are present on Langerhans cells, other cells of the dendritic cell family¹⁰¹ and monocytes, how their levels vary with maturation and activation, how readily these cells can be infected with various strains of HIV, and whether they can subsequently initiate infection of CD4⁺ cells. In addition to the deletions in the CCR5 gene already discussed, other less obvious mutations in CCR5, CXCR4 and other co-receptors may be identified that affect individual susceptibility to infection. Individual susceptibility to HIV-1 is very variable, however, and can vary with time within individuals; it is probably influenced not only by specific genetic defects, but also by the relative levels of different cell types, the activation of the cellular microenvironment, the patterns of cytokine expression and secretion, and the expression of co-receptors used by different viral strains. Expression of CD4 together with a given co-receptor does not necessarily render a target cell susceptible to the corresponding virus, as shown by monocyte/macrophages that express CXCR4 (ref. 60) but are resistant to T-tropic HIV-1 (ref. 62). Moreover, infection of monocyte/macrophages by M-tropic virus is unaffected by β -chemokines^{60,75,102}, suggesting that alternative entry mechanisms exist.

Progression of HIV-induced disease. The discovery of this complex array of host factors also has significant implications for our understanding of the transition from M- to T-tropic strains and the progression of disease in infected individuals⁶⁷ (Fig. 4). Despite the predominance of M-tropic strains in individuals with early or stable HIV infection, some T-tropic viruses are almost invariably present^{95,103}. It is still unclear why the M-tropic strains remain predominant and what triggers the eventual transition to T-tropic strains that precedes the drop in CD4⁺ T-cell counts and the development of AIDS^{66,67}. Were viral mutations alone responsible, the shift would probably occur rapidly in most infected persons, as T-tropic viruses are produced stochastically, and such viruses replicate more rapidly than M-tropic viruses *in vitro*¹³. The simplest model thus assumes that the dominance of M-tropic strains is maintained by a balance of conditions favouring the replication of M- rather than T-tropic strains, which must be disturbed by one or more events to trigger the transition. This is supported by the extended dominance of M-tropic viruses in many HIV-infected individuals, especially long-term non-progressors^{104,105}, which argues strongly that host factors are important in controlling the transition.

What these factors may be is a matter for speculation. As

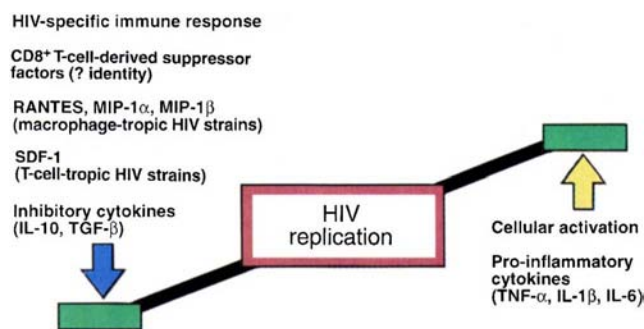


FIG. 4 Control of HIV replication and the progress of HIV-induced disease by the balance of host factors. Cellular activation and proinflammatory cytokines drive viral replication. These are counterbalanced by inhibitory factors, including both the immune system and nonspecific cytokines such as IL-10 and TGF- β , as well as co-receptor ligands such as the β -chemokines and SDF-1. Other inhibitory factors secreted by CD8⁺ T cells remain to be identified.

lymphoid tissues are the predominant sites of viral replication and the major viral reservoir^{1,8,106,107}, the lymphoid microenvironment almost certainly plays a critical role. Within these tissues, low expression of CXCR4 and/or high expression of SDF-1 would inhibit replication of T-tropic strains. Interestingly, SDF-1 is produced predominantly by stromal cells rather than lymphoid cells¹⁰⁸, and the normal stromal architecture of lymphoid tissue may initially favour high expression of this cytokine, which would be lost as the architecture breaks down. Disruption of the lymphoid tissue architecture is characteristic of progressive HIV disease⁸, and removal of a major inhibitor of T-tropic virus at the main site of viral replication could certainly initiate or exacerbate the shift to T-tropic strains. It will be important to examine SDF-1 levels in lymphoid tissue from various stages of HIV disease, as well as the effect of cytokines such as IFN- γ , TNF- α , IL-1 β and IL-6 (ref. 41) on the expression of co-receptors and their ligands.

As already mentioned, active infection with *M. tuberculosis* accelerates the course of HIV disease^{26,27} and the course of HIV infection in sub-Saharan Africa appears to be more fulminant than in developed countries^{28–31}, probably reflecting the persistent immune activation and high levels of cytokine secretion associated with chronic infection in both cases^{28,32}. In addition to markedly increasing HIV gene expression¹, persistent high levels of cellular activation and proinflammatory cytokine secretion may favour the emergence of T-tropic strains by accelerating viral replication (and hence mutation), as well as modulating the relative expression of co-receptors and their ligands. The profiles of HIV-co-receptors and their ligands in PMBCs and lymphoid tissue before and after infected individuals are immunized with common recall antigens should also be of interest.

It will also be important to determine whether mutations in the gene for CXCR4 or SDF-1 exist among HIV-infected individuals, particularly long-term non-progressors, which may interfere with their ability to propagate T-tropic viruses. Homozygous defects in human SDF-1 are unlikely to be found, as mice homozygous for deletions in the gene die perinatally¹⁰⁹, but it is unclear whether CXCR4 is likewise essential. Similarly, the lack of disease progression and low viral burden in chimpanzees infected with HIV-1 (ref. 110), as well as the lack of disease progression despite relatively high levels of virus in certain species of monkeys¹¹¹, may well be partly explained by the profile of viral co-receptors in these species.

Therapeutic and preventive strategies. Finally, appreciation of the role of host factors in HIV-induced pathogenesis suggests new therapeutic and preventive strategies. For therapy, it will be particularly important to confirm that individuals homozygous

for defects in the *CCR5* gene are phenotypically normal⁸³. CD4, the primary receptor for HIV-1, is essential for immune function¹¹² and interfering with its function for an extended period would be problematic. With *CCR5*, however, therapy could be aimed at blocking or downregulating *CCR5* using peptide or other ligands, or at interfering with the expression of the gene by gene therapy. For prevention, individuals could be vaccinated against co-receptor epitopes or against the domains of Env that interact with co-receptors^{90,91}, and it will be particularly important to map the latter precisely. The different co-receptors used by M- and T-tropic strains may also explain why vaccination with Env from T-tropic viruses induces neutralizing antibody responses against such strains, but not against primary isolates of HIV (ref. 113).

Applying these findings, however, may not be simple. The extraordinary redundancy and crossregulation of cytokines and their receptors³⁹ and the complexity of their interactions with HIV-1 mean that there is no guarantee that therapeutic or preventive interventions will be effective. Inhibitors of one receptor intended to block viral entry may upregulate other receptors and/or ligands, injuring a critical physiological function or even enhancing viral entry. Indeed, β -chemokines have recently been shown to enhance viral replication in monocyte/macrophages¹⁰². Therapeutic and preventive strategies affecting host factors, and particularly cytokines and their receptors, should be thus applied with considerable caution, beginning with animal models where possible. Despite these concerns, however, there can be no doubt that the new strategies suggested by recent advances in our understanding of host–HIV interactions have remarkable potential and should be vigorously pursued. □

Anthony S. Fauci is at the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

1. Fauci, A. S. *Science* **262**, 1011–1018 (1993).
2. Levy, J. A. *Microbiol. Rev.* **57**, 183–289 (1993).
3. Daar, E. S., Moudgil, T., Meyer, R. D. & Ho, D. D. *N. Engl. J. Med.* **324**, 961–964 (1991).
4. Clark, S. J. et al. *N. Engl. J. Med.* **324**, 954–960 (1991).
5. Moore, J. P., Cao, Y., Ho, D. D. & Koup, R. A. *J. Virol.* **68**, 5142–5155 (1994).
6. Koup, R. A. et al. *J. Virol.* **68**, 4650–4655 (1994).
7. Pantaleo, G., Graziosi, C. & Fauci, A. S. *N. Engl. J. Med.* **328**, 327–335 (1993).
8. Pantaleo, G. et al. *Nature* **362**, 355–358 (1993).
9. Piatak, M. et al. *Science* **259**, 1749–1754 (1993).
10. Ho, D. D. et al. *Nature* **373**, 123–126 (1995).
11. Wei, X. et al. *Nature* **373**, 117–122 (1995).
12. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. *Science* **271**, 1582–1586 (1996).
13. Coffin, J. M. *Science* **267**, 483–489 (1995).
14. Mellors, J. W. et al. *Science* **272**, 1167–1170 (1996).
15. Wolfs, T. F., Nara, P. L. & Goudsmit, J. *Chem. Immunol.* **56**, 1–33 (1993).
16. Haynes, B. F., Pantaleo, G. & Fauci, A. S. *Science* **271**, 324–328 (1996).
17. Zack, J. A. et al. *Cell* **61**, 213–222 (1990).
18. Ascher, M. S., Sheppard, H. W., Krowka, J. F. & Bremermann, H. J. *Adv. Exp. Med. Biol.* **374**, 203–210 (1995).
19. Bukrinsky, M. I., Stanwick, T. L., Dempsey, M. P. & Stevenson, M. *Science* **254**, 423–427 (1991).
20. Folks, T. et al. *J. Immunol.* **136**, 4049–4053 (1986).
21. Claydon, E. J., Bennett, J., Gor, D. & Forster, S. M. *AIDS* **5**, 113–114 (1991).
22. Goletti, D. et al. *J. Immunol.* **157**, 1271–1278 (1996).
23. Heng, M. C., Heng, S. Y. & Allen, S. G. *Lancet* **343**, 255–258 (1994).
24. Israel-Biet, D., Cadranet, J. & Even, P. *Am. Rev. Respir. Dis.* **148**, 1308–1312 (1993).
25. Denis, M. & Ghadirian, E. *Clin. Exp. Immunol.* **97**, 76–82 (1994).
26. Whalen, C. et al. *Am. J. Resp. Crit. Care Med.* **151**, 129–135 (1995).
27. Pape, J. W., Jean, S. S., Ho, J. L., Hafner, A. & Johnson, W. D. Jr. *Lancet* **342**, 268–272 (1993).
28. Quinn, T. C. et al. *J. Am. Med. Soc.* **257**, 2617–2621 (1987).
29. Nagelkerke, N. J. et al. *AIDS* **4**, 743–747 (1990).
30. Colebunders, R. L. & Latif, A. S. *AIDS* **5** (suppl. 1), S103–S112 (1991).
31. Wachter, H. et al. *Lancet* **2**, 1216–1217 (1986).
32. Bentwich, Z., Kalinkovich, A. & Weisman, Z. *Immunol. Today* **16**, 187–191 (1995).
33. Brown, C. C. et al. *J. Infect. Dis.* **169**, 975–980 (1994).
34. Thea, D. M. et al. *Ann. Intern. Med.* **124**, 757–762 (1996).
35. Ho, D. D. *Lancet* **339**, 1549 (1992).
36. Staprans, S. I. et al. *J. Exp. Med.* **182**, 1727–1737 (1995).
37. Stanley, S. K. et al. *N. Engl. J. Med.* **334**, 1222–1230 (1996).
38. O'Brien, W. A. et al. *Blood* **86**, 1082–1089 (1995).
39. Paul, W. E. & Seder, R. A. *Cell* **76**, 241–251 (1994).
40. Poli, G. & Fauci, A. S. in *Human Cytokines: Their Role in Disease and Therapy* (eds Aggarwal, B. B. & Puri, R. K.) 421–449 (Blackwell Science, Cambridge, MA, 1995).
41. Graziosi, C. et al. *Science* **265**, 248–252 (1994).
42. Mosmann, T. R. & Coffman, R. L. *Annu. Rev. Immunol.* **7**, 145–173 (1989).
43. Romagnani, S. *J. Clin. Immunol.* **15**, 121–129 (1995).
44. Clerici, M. & Shearer, G. M. *Immunol. Today* **14**, 107–111 (1993).
45. Clerici, M. & Shearer, G. M. *Immunol. Today* **15**, 575–581 (1994).
46. Maggi, E. M. *Science* **265**, 244–248 (1994).
47. Folks, T. M., Justement, J., Kinter, A., Dinarello, C. A. & Fauci, A. S. *Science* **238**, 800–802 (1987).

48. Kinter, A. L., Poli, G., Fox, L., Hardy, E. & Fauci, A. S. *J. Immunol.* **154**, 2448–2459 (1995).
49. Kinter, A. L. et al. *Proc. Natl Acad. Sci. USA* **93**, 14076–14081 (1996).
50. Weissman, D., Poli, G. & Fauci, A. S. *AIDS Res. Hum. Retr.* **10**, 1199–1206 (1994).
51. Lenardo, M. J. & Baltimore, D. *Cell* **58**, 227–229 (1989).
52. Duh, E. J., Maury, W. J., Folks, T. M., Fauci, A. S. & Rabson, A. B. *Proc. Natl Acad. Sci. USA* **86**, 5974–5978 (1989).
53. Poli, G. et al. *J. Exp. Med.* **172**, 151–158 (1990).
54. Walker, C. M., Moody, D. J., Stites, D. P. & Levy, J. A. *Science* **234**, 1563–1566 (1986).
55. Levy, J. A., Mackewicz, C. E. & Barker, E. *Immunol. Today* **17**, 217–224 (1996).
56. Barker, T. D., Weissman, D., Daucher, J. A., Roche, K. M. & Fauci, A. S. *J. Immunol.* **156**, 4476–4483 (1996).
57. Baier, M., Werner, A., Bannert, N., Metzner, K. & Kurth, R. *Nature* **378**, 563 (1995).
58. Cocchi, F. et al. *Science* **270**, 1811–1815 (1995).
59. Dragic, T. et al. *Nature* **381**, 667–673 (1996).
60. Moriuchi, H. et al. *Proc. Natl Acad. Sci. USA* (in the press).
61. Baggiolini, M., Dewald, B. & Moser, B. *Adv. Immunol.* **55**, 97–179 (1994).
62. Cheng-Mayer, C., Seto, D., Tateno, M. & Levy, J. A. *Science* **240**, 80–82 (1988).
63. Connor, R. I., Mohri, H., Cao, Y. & Ho, D. D. *J. Virol.* **67**, 1772–1777 (1993).
64. Roos, M. T. et al. *J. Infect. Dis.* **165**, 427–432 (1992).
65. Zhu, T. et al. *Science* **261**, 1179–1181 (1993).
66. Tersmette, M. et al. *Lancet* **(i)**, 983–985 (1989).
67. Schuitemaker, H. et al. *J. Virol.* **66**, 1354–1360 (1992).
68. Collman, R. et al. *J. Virol.* **66**, 7517–7521 (1992).
69. Shibata, R. et al. *J. Virol.* **69**, 4453–4462 (1995).
70. Doranz, B. J. et al. *Cell* **85**, 1149–1158 (1996).
71. Cheng-Mayer, C. *AIDS* **4** (suppl. 1), S49–S56 (1990).
72. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872–877 (1996).
73. Murphy, P. M. *Cytokine Growth Fact. Rev.* **7**, 47–64 (1996).
74. Choe, H. et al. *Cell* **85**, 1135–1148 (1996).
75. Deng, H. et al. *Nature* **381**, 661–666 (1996).
76. Alkhatib, G. et al. *Science* **272**, 1955–1958 (1996).
77. Bleul, C. C. et al. *Nature* **382**, 829–833 (1996).
78. Oberlin, E. et al. *Nature* **382**, 833–835 (1996).
79. Oravec, T., Pall, M. & Norcross, M. A. *J. Immunol.* **157**, 1329–1332 (1996).
80. Rowland-Jones, S. et al. *Nature Med.* **1**, 59–64 (1995).
81. Clerici, M. et al. *J. Infect. Dis.* **165**, 1012–1019 (1992).
82. Paxton, W. A. et al. *Nature Med.* **2**, 412–417 (1996).
83. Liu, R. et al. *Cell* **86**, 367–377 (1996).
84. Samson, M. et al. *Nature* **382**, 722–725 (1996).
85. Dean, M. et al. *Science* **273**, 1856–1861 (1996).
86. Huang, Y. et al. *Nature Med.* **2**, 1240–1243 (1996).
87. Weiss, R. A. & Clapham, P. R. *Nature* **381**, 647–648 (1996).
88. Bates, P. *Cell* **86**, 1–3 (1996).
89. Maddon, P. J. et al. *Cell* **47**, 333–348 (1986).
90. Wu, L. et al. *Nature* **384**, 179–183 (1996).
91. Trkola, A. et al. *Nature* **384**, 184–187 (1996).
92. Lapham, C. K. et al. *Science* **274**, 602–605 (1996).
93. Cocchi, F. et al. *Nature Med.* **2**, 1244–1247 (1996).
94. Hill, C. M. & Littman, D. R. *Nature* **382**, 668–669 (1996).
95. Cornelissen, M. et al. *J. Virol.* **69**, 1810–1818 (1995).
96. Loetscher, P., Seitz, M., Baggiolini, M. & Moser, B. *J. Exp. Med.* **184**, 569–577 (1996).
97. Moore, J. P. & Koup, R. A. *J. Exp. Med.* **184**, 311–313 (1996).
98. Spijkerman, I. J. et al. *AIDS* **9**, 1085–1092 (1995).
99. van't Wout, A. B. et al. *J. Clin. Invest.* **94**, 2060–2067 (1994).
100. Spira, A. I. et al. *J. Exp. Med.* **183**, 215–225 (1996).
101. Steinman, R. M. *Annu. Rev. Immunol.* **9**, 271–296 (1991).
102. Schmidtmayerova, H., Sherry, B. & Bukrinsky, M. *Nature* **382**, 767 (1996).
103. Valentin, A., Albert, J., Fenyo, E. M. & Asjo, B. *J. Virol.* **68**, 6684–6689 (1994).
104. Pantaleo, G. et al. *N. Engl. J. Med.* **332**, 209–216 (1995).
105. Cao, Y., Qin, L., Zhang, L., Safrit, J. & Ho, D. D. *N. Engl. J. Med.* **332**, 201–208 (1995).
106. Embretson, J. et al. *Nature* **362**, 359–362 (1993).
107. Pantaleo, G. et al. *Immunol. Rev.* **140**, 105–130 (1994).
108. Nagasawa, T., Kikutani, H. & Kishimoto, T. *Proc. Natl Acad. Sci. USA* **91**, 2305–2309 (1994).
109. Nagasawa, T. et al. *Nature* **382**, 635–638 (1996).
110. Fultz, P. N. *Clin. Infect. Dis.* **17** (suppl. 1), S230–S235 (1993).
111. Desrosiers, R. C. *Annu. Rev. Immunol.* **8**, 557–578 (1990).
112. Rahemtulla, A. et al. *Nature* **353**, 180–184 (1991).
113. Mascola, J. R. et al. *J. Infect. Dis.* **173**, 340–348 (1996).

ACKNOWLEDGEMENTS. I thank G. Folkers, D. Weissman, A. Kinter, G. Scala, J. Arthos, O. Cohen, H. Moriuchi, M. Ostrowski and J. F. Dickson III for helpful discussions and for reviewing the manuscript.

KNOW YOUR COPY RIGHTS RESPECT OURS

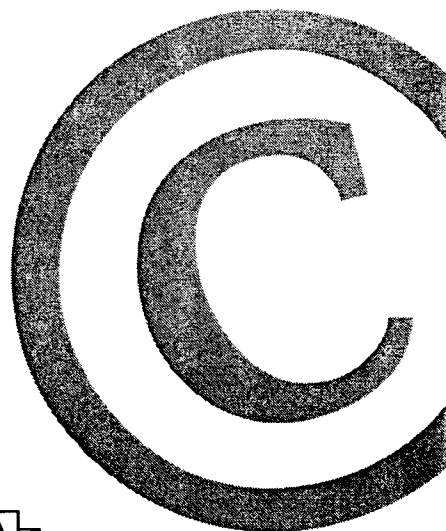
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

Evidence for a magnetosphere at Ganymede from plasma-wave observations by the Galileo spacecraft

D. A. Gurnett*, W. S. Kurth*, A. Roux†, S. J. Bolton‡ & C. F. Kennel§

* Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA

† Centre d'Etudes des Environnements Terrestre et Planétaires, Université Versailles Saint Quentin, 10/12 Avenue de l'Europe, 78140 Velizy, France

‡ Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

§ Office of the Chancellor, UCLA, Los Angeles, California 90095, USA

ON 27 JUNE 1996 the Galileo spacecraft^{1,2} made the first of four planned close fly-bys of Ganymede, Jupiter's largest moon. Here we report measurements of plasma waves and radio emissions, over the frequency range 5 Hz to 5.6 MHz during the first encounter. Intense plasma waves were detected over a region of space nearly four times Ganymede's diameter, which is much larger than would be expected for a simple wake arising from Ganymede's passage through Jupiter's rapidly rotating magnetosphere. The types of waves detected (whistler-mode emissions, upper hybrid waves, electrostatic electron cyclotron waves and escaping radio emission) strongly suggest that Ganymede has a large, extended magnetosphere of its own. The data indicate the presence of a strong ($B > 400$ nT) magnetic field, and show that Ganymede is surrounded by an ionosphere-like plasma with a maximum electron density of about 100 particles cm^{-3} and a scale height of about 1,000 km.

The trajectory for the Ganymede fly-by is shown in Fig. 1. This fly-by was designed to obtain measurements in the wake created by the co-rotating magnetospheric plasma of Jupiter. The closest approach occurred at 06:29:07 UT (universal time, h:min:s, at the spacecraft) at an altitude of 835 km and a radial distance of 1.32 times the radius of Ganymede, $1.32 R_G$ (where $1 R_G = 2,634$ km). Spectrograms of the electric and magnetic field intensities obtained from the plasma-wave instrument for a 2.5-hour interval around the time of closest approach are shown in Fig. 2. The intensities are colour coded with red being the most intense and blue being the least intense. A major response is evident for a period of nearly 50 minutes (from 06:14 to 07:02 UT) around the time of closest approach. This response lasts much longer than the time it takes to cross the diameter of Ganymede, which is only about 12 minutes. As no other comparable response has been observed elsewhere in Galileo's orbit, it is clear that these effects are associated with Ganymede.

Because the magnetic field measurements are the simplest to analyse, we start by discussing the magnetic field intensities (Fig. 2b). The onset of the magnetic field response associated with Ganymede is somewhat confused by interference from the ultraviolet spectrometer instrument from about 06:13 to 06:18 UT (labelled 'interference' in Fig. 2b). A few minutes after this interference ends, at about 06:23 UT, a broad irregular band of noise can be seen starting in the frequency range around a few hundred hertz. This noise is associated with Ganymede. Over the next few minutes, this band of noise becomes more intense and gradually rises in frequency, eventually reaching a peak frequency of about 6 kHz at about 06:29 UT, near closest approach. After closest approach the noise gradually declines in frequency and continues with relatively high intensities for about 30 minutes, to about 06:58 UT. A similar band of noise is also present in the

electric field spectrogram (Fig. 2a).

As the band of noise described above has both electric and magnetic fields, it clearly consists of electromagnetic waves. In this frequency range, well above the ion cyclotron frequencies and (as we will show) below the electron plasma frequency, the only electromagnetic plasma wave mode that can propagate is the whistler mode³. The presence of strong whistler-mode emissions in the vicinity of Ganymede is surprising, as these emissions are normally generated by energetic radiation belt electrons⁴. Although the spacecraft is located in the jovian radiation belt, there is no evidence of comparable whistler-mode emissions either before or after the Ganymede fly-by, so the emissions are clearly associated with Ganymede. Fine structure in the spectrum also shows that the emissions are a mixture of 'chorus' and 'hiss' (these types of whistler-mode emissions are defined in ref. 5). Chorus and hiss are commonly observed in the magnetospheres of the magnetized planets⁶.

The presence of whistler-mode emissions has important implications concerning the magnetic field strength in the vicinity of Ganymede. It is known³ that the whistler mode cannot propagate at frequencies above the electron cyclotron frequency, $f_c = 28 B$ Hz, where B is the magnetic field strength in nanotesla. In fact, whistler-mode emissions are seldom observed at frequencies above one-half the electron cyclotron frequency⁷. As the maximum frequency observed near closest approach is approximately 6 kHz, this implies that the electron cyclotron frequency must be at least 12 kHz. From the equation $f_c = 28 B$, it follows that the magnetic field strength must be at least 400 nT. This field strength is considerably stronger than the jovian magnetic field at the orbit of Ganymede which is only about 100 nT. We conclude that a relatively strong magnetic field must exist around Ganymede. This result is in agreement with the on-board magnetometer measurements of Kivelson *et al.*⁸, which show that Ganymede has an internally generated magnetic field with an equatorial surface field strength of about 750 nT. The electron cyclotron frequency computed from the magnetometer data is shown by the white line labelled f_c in Fig. 2b. As can be seen, the maximum frequency of the whistler-mode emissions is approximately one-half the electron cyclotron frequency.

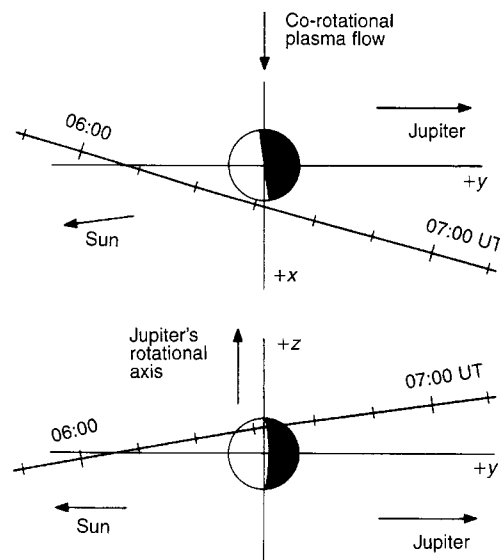


FIG. 1 The trajectory of Galileo during the first Ganymede fly-by. The Ganymede-centred coordinate system has the $+z$ axis aligned parallel to Jupiter's rotational axis and the $+x$ axis parallel to the nominal co-rotational plasma flow induced by Jupiter's rotation. The top diagram shows the view looking towards the $+z$ axis, and the bottom diagram shows the view looking towards the $+x$ axis.

We now consider the electric field spectrum. On the inbound pass, a series of broadband electric field noise bursts can be seen in Fig. 2a, from about 06:14 to 06:20 UT, and again on the outbound pass, from about 06:58 to 07:02 UT. These broadband noise bursts effectively define the outer limits of the plasma wave response associated with Ganymede. The noise bursts are most intense at low frequencies, but extend with measurable intensities to frequencies as high as 10^5 Hz. No comparable response can be seen in the magnetic field, so the noise is electrostatic. Broadband electrostatic noise bursts of this type are commonly observed in planetary magnetospheres, and typically occur at boundaries, such as the bow shock⁹ and magnetopause¹⁰. Comparisons with the magnetometer data of Kivelson *et al.*⁸ show that the innermost noise burst at 06:19 UT on the inbound pass, and a similar noise burst at 07:00 UT on the outbound pass, coincide with abrupt magnetic field rotations that they identify as magnetopause crossings. Whether some of the remaining noise bursts could be associated with a bow shock remains to be investigated.

Immediately after the inbound magnetopause crossing, a narrowband emission can be seen in the electric field spectrogram starting at a frequency of about 20 kHz at 06:20 UT, rising to a frequency of about 60 kHz at 06:30 UT, and then declining to about 16 kHz at 06:40 UT. Narrowband emissions of this type are a common feature of planetary magnetospheres^{11–14}, and are caused by electrostatic waves at the upper hybrid resonance frequency, f_{UH} . Upper hybrid emissions can be used to provide very accurate measurements of the local electron density. The upper hybrid frequency is given by $f_{UH} = (f_p^2 + f_c^2)^{1/2}$, where $f_p = 8,980\sqrt{N}$ Hz is the electron plasma frequency and N is the electron number density in cm^{-3} (ref. 3). From these equa-

tions, one can show that the electron density is given by $N = (f_{UH}^2 - f_c^2)/(8,980)^2 \text{ cm}^{-3}$. Because $f_{UH} \gg f_c$, the cyclotron frequency contribution can be ignored. The electron density computed using this technique is shown in Fig. 3b. As can be seen, the electron density profile has a parabolic shape that is nearly symmetric around closest approach, starting at about 5 cm^{-3} at 06:20 UT, rising to about 45 cm^{-3} at 06:30 UT, and declining to about 4 cm^{-3} at 06:40 UT. A corresponding radial profile is shown in Fig. 3a. The close similarity between the inbound and outbound radial profiles shows that Ganymede is surrounded by a dense plasma envelope with a scale height of about 1,000 km. If the observed height dependence is extrapolated downwards, the electron density near the surface is estimated to be about 100 cm^{-3} . Such large electron densities imply a substantial plasma source near Ganymede, probably caused by ionization of a neutral gas. At present, the only gaseous molecule known to exist at Ganymede is ozone, which is trapped in the surface ice¹⁵. These results strongly imply the existence of a substantial atmosphere around Ganymede.

During the interval from about 06:40 to 07:00 UT, a series of narrowband emissions can be seen from about 5 to 20 kHz. Emissions of this type are commonly observed in planetary magnetospheres and are called electrostatic electron cyclotron waves^{16–18}. Electrostatic electron cyclotron waves occur near harmonics of the electron cyclotron frequency and are generated by highly anisotropic electron velocity distributions of the type often found in planetary radiation belts. Careful inspection of the electric field spectrogram also clearly shows a weak band of radio emissions propagating outward from the region where the upper hybrid waves are observed (labelled 'radio emission' in Fig. 2). These radio emissions are believed to be produced by a process called mode conversion, wherein the electrostatic energy associated with the upper hybrid waves is converted to escaping electromagnetic radiation. Radio emissions produced by such a mode conversion process are a common feature of planetary magnetospheres^{19–22}.

We now discuss in more detail the mechanisms by which the waves observed near Ganymede are generated. The whistler-mode emissions are of particular interest. According to Kennel and Petschek⁴, the essential feature required to produce whistler-mode emissions is the presence of energetic electrons with a loss-cone anisotropy. A loss-cone anisotropy is produced when particles moving within a cone of directions along the magnetic field (the loss cone) strike the planet and are lost from the system. Such loss-cone distributions are normally associated with radiation belt particles trapped on 'closed' magnetic field lines that link opposite magnetic poles of the same body. However, according to the magnetic field model of Kivelson *et al.*⁸, Galileo was on 'open' magnetic field lines during the entire fly-by, that is, field lines that link the magnetic field of Ganymede to the magnetic field of Jupiter. This unique situation arises because Ganymede is located within the magnetosphere of Jupiter. Therefore, it does not appear that the whistler-mode emissions are produced by electrons trapped in a radiation belt at Ganymede. Absorption by Ganymede probably simply imposes a loss cone on the pre-existing energetic jovian electron population, thereby leading to the growth of whistler-mode waves. A rough estimate can be made of the electron energies responsible for the whistler-mode emissions. Kennel and Petschek⁴ show that the parallel energy of electrons in cyclotron resonance with a whistler-mode wave of frequency $\omega = 2\pi f$ is given by

$$W_{\parallel} = \frac{B^2}{8\pi N} \frac{\omega_c}{\omega} \left(1 - \frac{\omega}{\omega_c}\right)^3$$

where $\omega_c = 2\pi f_c$. Using parameters representative of the region near closest approach ($B \approx 500$ nT and $N \approx 45 \text{ cm}^{-3}$), one can show that $B^2/(8\pi N)$ is about 14 keV. For a wave frequency of 1 kHz, the resonant energy is then about 150 keV. Intense fluxes of electrons with energies in this range are known to be present in Jupiter's magnetosphere^{23–25}. The enhanced electron densities in

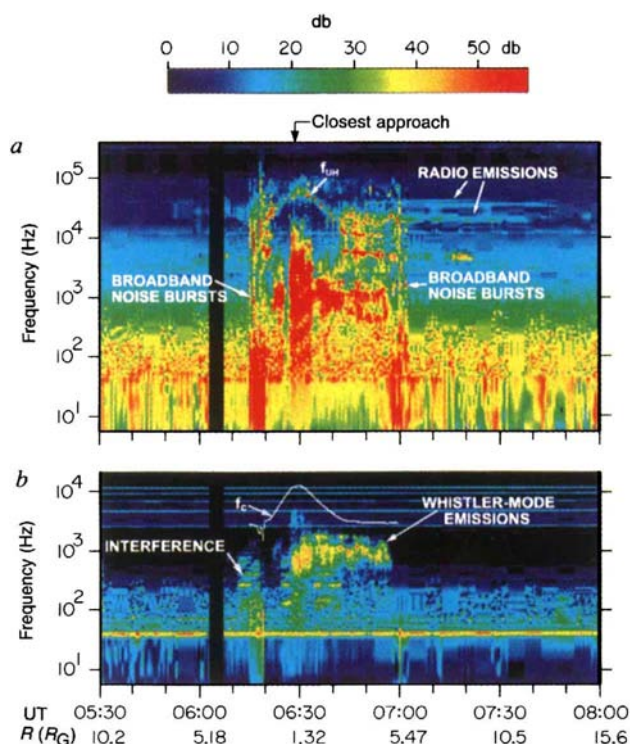


FIG. 2 Frequency-time spectrograms of the electric (a) and magnetic (b) field intensities detected by the Galileo plasma wave instrument during the 27 June 1996 Ganymede fly-by. The radial distance from the centre of Ganymede, R , is given in Ganymede radii, R_G . The time of closest approach is indicated at the top. The electron cyclotron frequency, f_c , shown by the white line in b, was computed from the magnetometer data of Kivelson *et al.*⁸. Various horizontal lines, particularly in the magnetic field spectrogram, are caused by spacecraft-generated interference. The colour scale shows the intensity in decibels.

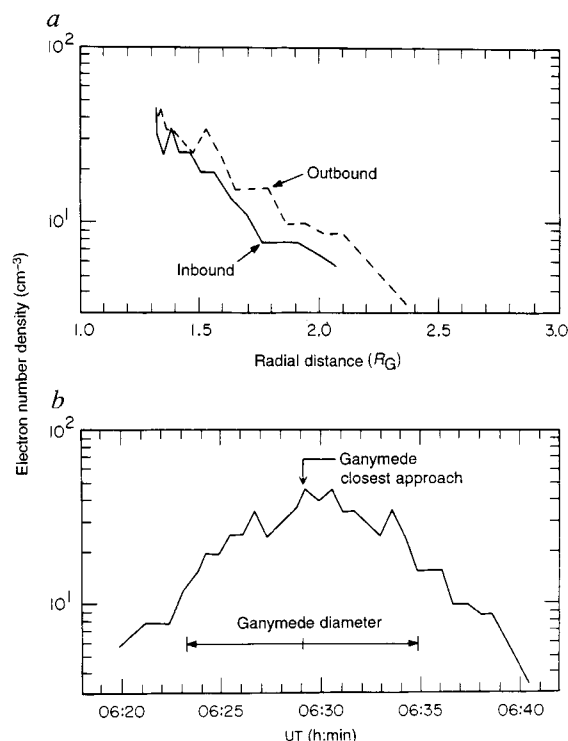


FIG. 3 The lower panel shows the electron density (N) determined from upper hybrid emissions in the vicinity of Ganymede. The upper panel shows the electron number density as a function of radial distance (R). Note the close similarity of the inbound and outbound profiles, which indicates that Ganymede is surrounded by an ionosphere-like plasma envelope extending several thousand kilometres above the surface.

the vicinity of Ganymede probably play an important role in the generation of these emissions by bringing the resonance energy down into a range where there are high electron intensities.

The electrostatic upper hybrid waves and electron cyclotron waves are closely related and constitute a general class known as electron cyclotron harmonic (ECH) waves. ECH waves are also driven by an anisotropy in the electron velocity distribution, similar to the whistler-mode. However, the electron energies involved¹⁸, 10–100 eV, are usually much lower, and the origin of the anisotropy is quite different. Whereas absorption by a planetary body usually causes the anisotropy required to drive whistler-mode emissions, other factors usually cause the anisotropy required to produce electron cyclotron waves. For example, when electrons are convected into a region of strong magnetic field, the perpendicular energy is increased relative to the parallel energy ($W_{\perp} > W_{\parallel}$), thereby producing the necessary anisotropy. The electron plasma frequency to cyclotron frequency ratio, f_p/f_c , also plays a crucial role¹⁷. This ratio varies considerably during the fly-by, which may explain why the whistler-mode waves and the electron cyclotron waves occur in different regions. Electron precipitation into Ganymede's atmosphere due to pitch-angle scattering by whistler-mode or ECH waves could cause observable optical emissions, possibly accounting for the auroral emissions recently observed near Ganymede by the Hubble Space Telescope²⁶. □

Received 30 September; accepted 5 November 1996.

1. Johnson, T. V., Yeates, C. M. & Young, R. *Space Sci. Rev.* **60**, 3–21 (1992).
2. Gurnett, D. A. et al. *Space Sci. Rev.* **60**, 341–355 (1992).
3. Stix, T. H. *The Theory of Plasma Waves* 12 (McGraw-Hill, New York, 1962).
4. Kennel, C. F. & Petschek, H. E. *J. Geophys. Res.* **71**, 1–28 (1966).
5. Helliwell, R. A. *Whistlers and Related Ionospheric Phenomena* 207 (Stanford Univ. Press, Stanford, 1965).

6. Kurth, W. S. & Gurnett, D. A. *J. Geophys. Res.* **96**, 18977–18991 (1991).
7. Burtis, W. J. & Helliwell, R. A. *J. Geophys. Res.* **74**, 3002–3010 (1969).
8. Kivelson, M. G. et al. *Nature* **384**, 537–541 (1996).
9. Scarf, F. L., Gurnett, D. A. & Kurth, W. S. *Nature* **292**, 747–750 (1981).
10. Gurnett, D. A. et al. *J. Geophys. Res.* **84**, 7043–7058 (1979).
11. Walsh, D., Haddock, T. F. & Schulte, H. F. *Space Res.* **4**, 935–959 (1964).
12. Mosier, S. R., Kaiser, M. L. & Brown, L. W. *J. Geophys. Res.* **78**, 1673–1677 (1973).
13. Warwick, J. W. et al. *Science* **204**, 995–998 (1979).
14. Gurnett, D. A. *J. Geophys. Res.* **86**, 8199–8212 (1981).
15. Noll, K. S., Johnson, R. E., Lane, A. L., Domingua, D. L. & Weaver, H. A. *Science* **273**, 341–343 (1996).
16. Kennel, C. F., Scarf, F. L., Fredericks, R. W., McGehee, J. H. & Coroniti, F. V. *J. Geophys. Res.* **75**, 6136–6152 (1970).
17. Ashour-Abdalla, M., Chanteur, G. & Pellat, R. *J. Geophys. Res.* **80**, 2775–2782 (1975).
18. Rönnmark, K., Borg, H., Christiansen, P. J., Gough, M. P. & Jones, D. *Space Sci. Rev.* **22**, 401–417 (1978).
19. Gurnett, D. A. *J. Geophys. Res.* **80**, 2751–2763 (1975).
20. Gurnett, D. A. & Frank, L. A. *J. Geophys. Res.* **81**, 3875–3885 (1976).
21. Kaiser, M. L. & Desch, M. D. *J. Geophys. Res.* **87**, 389–392 (1980).
22. Melrose, D. B. *J. Geophys. Res.* **86**, 30–36 (1981).
23. Van Allen, J. A. et al. *Science* **183**, 309–311 (1974).
24. Krimigis, S. M. et al. *Science* **204**, 998–1003 (1979).
25. Lanzerotti, L. J. et al. *Science* **257**, 1518–1524 (1992).
26. Cowen, R. *Science News* **150**, 181 (1996).

ACKNOWLEDGEMENTS. We thank the Galileo project team at the Jet Propulsion Laboratory for their efforts in obtaining these data, and M. Kivelson for providing the magnetometer data. We also thank L. Granroth and J. Groene for their assistance in the data processing at The University of Iowa. This work was supported by NASA through the Jet Propulsion Laboratory, and by the Centre National d'Etudes Spatiales (France).

CORRESPONDENCE should be addressed to D.A.G. (donald-gurnett@uiowa.edu).

Discovery of Ganymede's magnetic field by the Galileo spacecraft

M. G. Kivelson^{*†}, K. K. Khurana^{*}, C. T. Russell^{*†},
R. J. Walker^{*}, J. Warnecke^{*}, F. V. Coroniti[‡],
C. Polanskey[§], D. J. Southwood^{*||} & G. Schubert^{*‡}

^{*} Institute of Geophysics and Planetary Physics, [†] Department of Earth and Space Sciences, [‡] Department of Physics, University of California, Los Angeles, California 90095-1567, USA

[§] Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

^{||} Department of Physics, Imperial College of Science, Technology, and Medicine, London SW7 2BZ, UK

THE Galileo spacecraft has now passed close to Jupiter's largest moon—Ganymede—on two occasions, the first at an altitude of 838 km, and the second at an altitude of just 264 km. Here we report the discovery during these encounters of an internal magnetic field associated with Ganymede (the only other solid bodies in the Solar System known to have magnetic fields are Mercury, Earth and probably Io¹). The data are consistent with a Ganymede-centred magnetic dipole tilted by $\sim 10^\circ$ relative to the spin axis, and an equatorial surface-field strength of ~ 750 nT. The magnetic field is strong enough to carve out a magnetosphere with clearly defined boundaries within Jupiter's magnetosphere. Although the observations require an internal field, they do not indicate its source. But the existence of an internal magnetic field should in itself help constrain models of Ganymede's interior.

On Galileo's first inbound pass following orbital insertion, the magnetometer² measurements followed reasonably closely the predictions from a recent model of the magnetic field of Jupiter's magnetosphere³ that we refer to as the KK96 model. (This model consists of the O6 model⁴ of Jupiter's internal field plus the field of a warped and hinged current sheet parametrized to fit the magnetic field measured on the Pioneer 10 outbound pass near the dawn meridian.) The field increased in magnitude with approach to Jupiter and varied in orientation at Jupiter's rotation period. Data from 00:00 UT (universal time; h:min) to 12:00 UT on

27 June 1996 are plotted in Fig. 1; the Ganymede-associated perturbation is clearly apparent at 06:29 UT. The Ganymede encounter occurred well off the jovian magnetic equator (identifiable by the reversal of sign of B_r (the coordinate system is defined in Fig. 1 legend)) and thus in a region of relatively low plasma density. Closest approach was downstream of Ganymede relative to the flowing plasma of Jupiter's magnetosphere. Near to Ganymede, the field increased from 107 nT at 06:00 UT to 480 nT at 06:29:07 UT (h:min:s) near closest approach, and then decreased to background, returning to 118 nT at ~07:00 UT. A considerable rotation accompanied the change in field magnitude.

The data near closest approach for this first pass are shown in Fig. 2. The field near closest approach is approximately that of a magnetic dipole centred at Ganymede with surface strength 750 nT at the equator added to the KK96 model. The dipole north pole is tilted by 10° from the spin axis towards 200° Ganymede east longitude. (Longitude of 180° is radially outward from Jupiter.) Currents flowing in the magnetospheric plasma in Ganymede's environment perturb the magnetic signature^{5,6}, so the actual magnetic moment of Ganymede may be slightly overestimated by the vacuum field estimate. In Fig. 3 we show projections of the trajectory and of the perturbation field vectors measured along it. (The perturbation field is the vector difference between the observed field and the KK96 model field.) The figure illustrates that the field perturbations converge towards Ganymede, as expected in the vicinity of an internal dipole.

For the second pass, Ganymede was again located well above the jovian current sheet, probably in a region of low plasma density. The full data from the magnetometer and other particle and field instruments will be reported elsewhere, but measurements at ~1-minute resolution were acquired in advance of the full data set by operating the magnetometer² in a mode that averages and stores up to 200 field vectors for delayed transmission to the ground. Figure 4 presents these initial data and the vacuum field model previously described, which again provides a reasonable approximation to the data. Because of the low altitude of closest approach on this pass, the internal field dominates the signature (the field magnitude reached 1,146 nT at 18:59:45 on 6 September 1996). The small discrepancies between the data and the modelled field arise from various neglected effects. The vacuum superposition model does not allow for the fact that the external field is frozen into the flowing plasma of the jovian magnetosphere which modifies the interaction. In particular, currents that flow through the jovian plasma and close through Ganymede or a possible ionosphere^{5,6} produce perturbations that tilt the field in the direction of the corotation flow. These neglected perturbations affect principally B_ϕ (see Fig. 2 legend) which is poorly fitted by the model. Possible contributions from higher-order multipoles of the internal field have not been considered.

We have traced field lines near Ganymede in a simplified representation of the magnetic geometry (Fig. 5). The background (jovian) field is assumed uniform and tilted radially away from Jupiter. It has been added to the field of a Ganymede dipole pointing to 10° radially inward relative to the southward direction (which is the same as tilting the

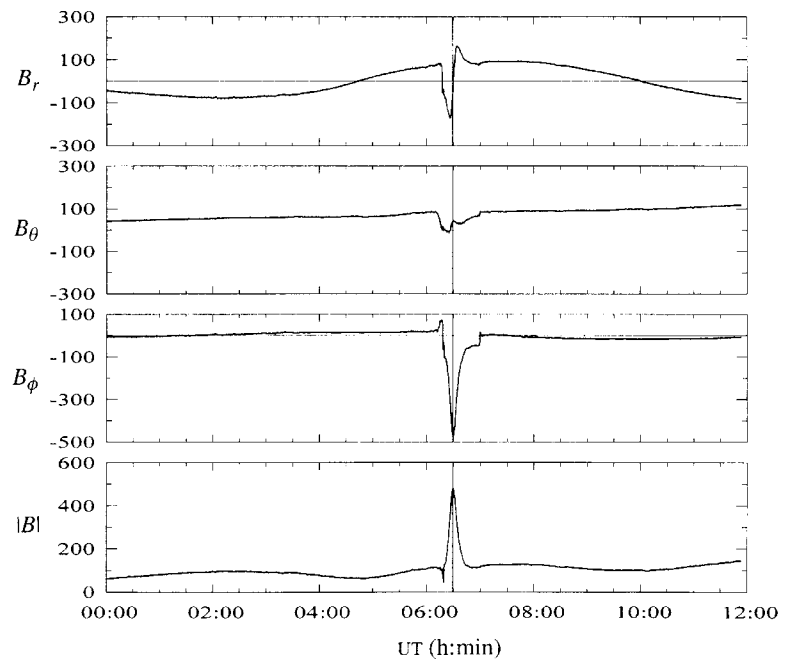


FIG. 1 Magnetometer data for the inbound pass to Jupiter on 27 June 1996 over a radial distance range from 17.44 to 13.14 jovian radii from Jupiter. The data are given in a Jupiter-centred right-handed spherical coordinate system (r, θ, ϕ) with r the radial distance, θ the angle from the spin axis, and ϕ the azimuthal angle. This is a variant of System III (1996)²⁴ with ϕ increasing westward. The ten-hour periodicity arises from Jupiter's rotation. Closest approach to Ganymede was at 06:29:07 UT. All times are given as spacecraft event time, which is universal time at the spacecraft. The magnetic field vector is $\mathbf{B} = (B_r, B_\theta, B_\phi)$.

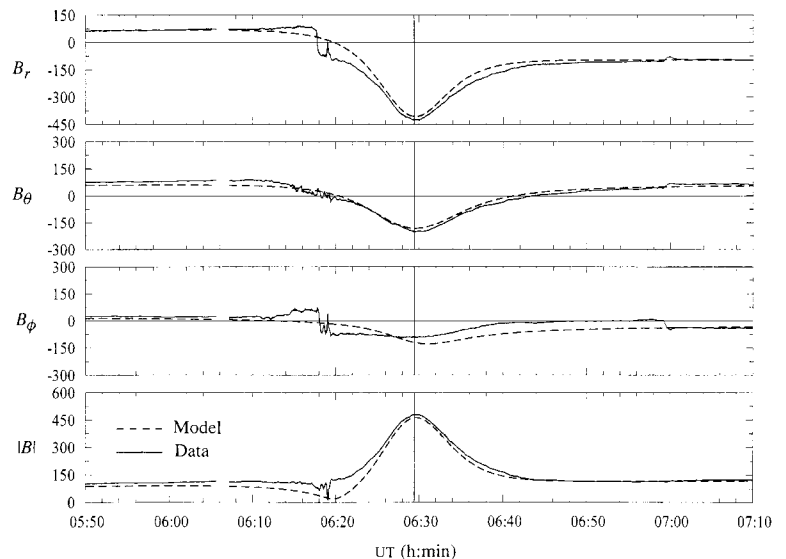


FIG. 2 Magnetic-field components and magnitude (solid lines) from 05:50 UT to 07:10 UT on 27 June 1996. The data are given in a Ganymede-centred coordinate system referenced to Ganymede's spin axis with r the radial distance, θ the angle from an axis parallel to Jupiter's spin axis and through Ganymede's centre, and ϕ the azimuthal angle, increasing eastward. Galileo's radial distance from the centre of Ganymede varied from $6.85 R_G$ (radius of Ganymede, 2,634 km) at 05:50 UT to $1.32 R_G$ at closest approach (marked by a vertical line) and out to $5.47 R_G$ at 07:00 UT. Dashed lines show the superposition of a model of the jovian field³ and the field of a Ganymede-centred magnetic moment described in the text. The data resolution is 0.333 s between 06:07 and 06:44 UT. Elsewhere the data resolution is 24 s.

north pole 10° radially outward). This model neglects minor components of both the external field and the dipole moment out of the y - z plane (the coordinates are defined in Fig. 3 legend). A significant element of the schematic is the presence of field lines linked to Ganymede at both ends out to $\sim 2R_G$ (where R_G is the radius of Ganymede), implying that Ganymede carves out its own magnetosphere within the jovian magnetosphere. The interpretations of plasma wave emissions by Gurnett⁷ are consistent with this interpretation.

Figure 5 shows (heavy lines) the separatrix that encloses Ganymede's magnetosphere, identified as the region in which field lines link to Ganymede at least once. Galileo crossed into the Ganymede magnetosphere very near the boundary between field lines that intersect Ganymede's surface only once and field lines that intersect Ganymede's surface twice (see the trajectory in Fig. 5) and then passed through a region analogous to the cusp (or polar cusp) of a conventional magnetosphere⁸. The separatrix intersects the inbound trajectory at 06:20 UT, quite close to the time of the rather abrupt field rotation observed at $\sim 06:17:45$ UT, and intersects the outbound trajectory at 07:00 UT just at the time a small rotation was observed. The boundary is analogous to the magnetopause of a conventional magnetosphere, and its signature is also clear in the plasma wave data⁷. The trajectory passed through the magnetic neutral line in the schematic simultaneous with crossing of the inbound separatrix. In the data, the field magnitude dropped to 12.85 nT consistent with a near-neutral line encounter at 06:19:06 UT, more than one minute after the field rotation.

We have analysed the field rotation for the inbound magnetopause crossing. The normal to the surface lies almost in the y - z plane of Fig. 5a and is rotated radially outwards from the z -axis by 33° . This orientation suggests that in a more realistic model, the dipole centre might be shifted slightly towards positive z , the dimple in the separatrix would shift upward, and the trajectory would cross the separatrix just below the dimple. The small shift would produce a delay between the crossing of the separatrix and the encounter with the neutral line. It would account for the orientation of the magnetopause boundary during the inbound crossing, and it would hardly affect the outbound crossing. (Because the rotation at the outbound magnetopause is small, we are not able to evaluate the normal from the low-resolution data.)

It is not clear that we are justified in taking the details of a vacuum superposition model as far has been done in the above analysis, but, because the ambient jovian electron density was probably $\leq 1 \text{ cm}^{-3}$, it is likely that neither the Alfvénic Mach number (flow speed/Alfvén speed) nor the plasma β (thermal pressure/magnetic pressure) was large enough for plasma effects to displace the boundary significantly. However, the Ganymede magnetosphere must divert the jovian plasma around the $\sim 4R_G$ (diameter) magnetic obstacle. This region of diverted flow is analogous to a traditional magnetosheath, although the diverted flow is not bounded by a shock. Just before the magnetopause crossing, low-frequency fluctuations were observed as is common in a magnetosheath. Elsewhere fluctuations are very small, although the time resolution of the data was sufficient to detect fluctuations on the scale of seconds between 06:07 UT and 06:44 UT. The absence of fluctuations suggests low plasma pressure within the magnetosphere, although a peak density of $>45 \text{ cm}^{-3}$ has been reported⁷.

The dipole moment of Ganymede ($1.4 \times 10^{13} \text{ T m}^3$ or $1.4 \times 10^{20} \text{ A m}^2$) is close to that inferred for Io^{1,9}. Thus Ganymede, whose angular momentum is also close to that of Io, follows the trend of the relation between the magnetic moment and the angular momentum for magnetized planets suggested by Blackett¹⁰. However, the discovery of an internal field at Ganymede does not bear directly on the interpretation of the signature at Io as the properties of the two bodies differ greatly. In some ways, the conditions at Io seem more favourable for field generation than those at Ganymede. Density is higher, a large core is known to be

present¹¹, rotation is faster, and there is a source of heat that could drive convection, even though the heat is not deposited directly in the core. However, as details of the heating and transport processes are likely to be very different at the two jovian satellites, the discovery of a Ganymede field does little to support the arguments^{1,9} for an internal magnetic field of Io.

Given that the dipole moments are comparable, it may seem puzzling that the evidence for an internal field at Ganymede is unambiguous whereas the evidence for an internal field at Io, though compelling, is not unambiguous. The ambiguity arises both because the ratio of the magnitudes of the nominal surface equatorial magnetic field to the ambient jovian magnetic field is smaller by almost an order of magnitude at Io than at Ganymede and because the signature near Io is partially obscured by strong perturbations from ion pickup and charge exchange.

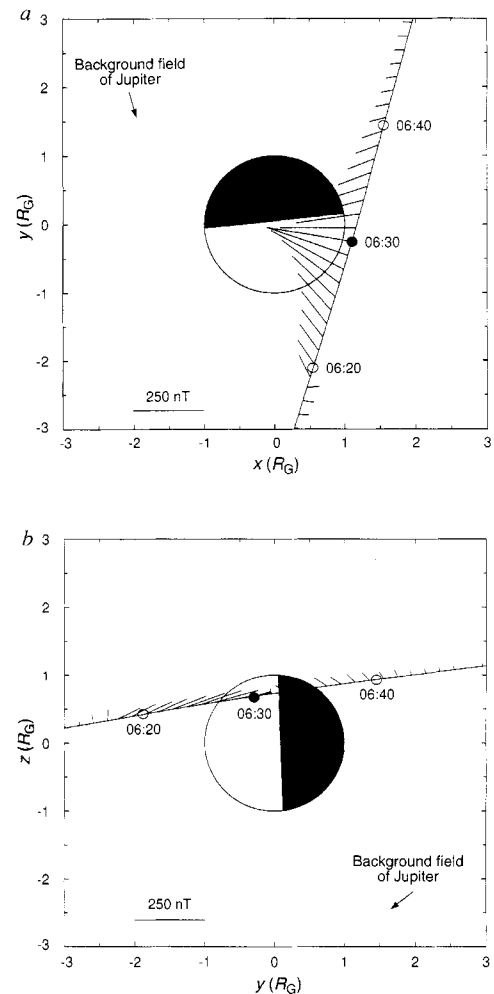
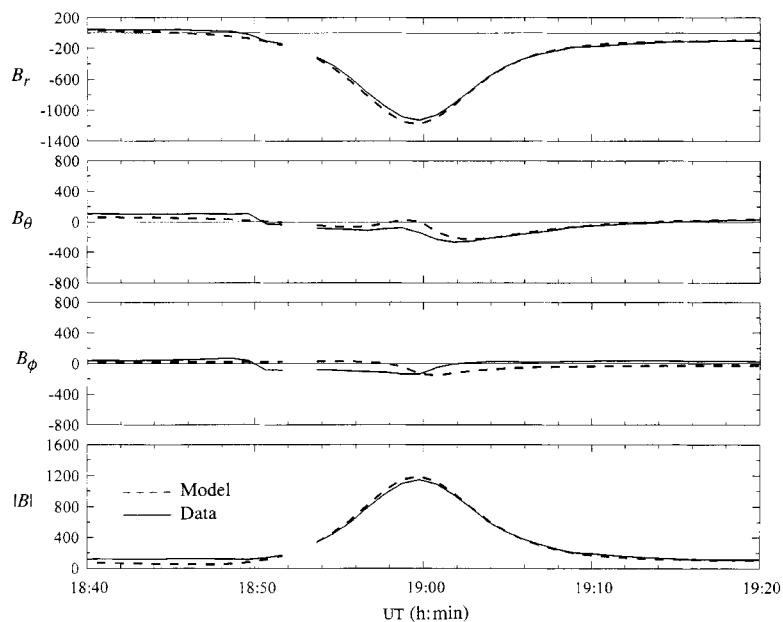


FIG. 3 Galileo's (27 June 1996) inbound towards Jupiter in the region near Ganymede (large circle, Ganymede; small circles, Galileo). The filled circle is near the closest approach to Ganymede. The plots use a coordinate system referenced to the direction of co-rotation (along $\hat{x}$) and the spin axis of Jupiter, effectively the spin axis of Ganymede (along $\hat{z}$). The unperturbed background field lies in the y - z plane roughly 50° outward from $-\hat{z}$; $\hat{y}$ is positive inward towards Jupiter. a, x - y projection, indicating the flow direction. The lines rooted along the trajectory are proportional to the projection of $\mathbf{B} - \mathbf{B}_{\text{model}}$ of Fig. 2 and the scale for the field perturbations is indicated. Key times are given. The projection direction of the background field is also shown. The right side of Ganymede is shaded. The terminator was crossed close to the centre of the wake, with the sunlit side corresponding to negative values of y . b, y - z projection of the trajectory and the perturbation field vectors. Note that the trajectory passes principally above Ganymede's equator.

FIG. 4 Magnetic field components and magnitude (solid lines) from 18:40 to 19:20 UT on 6 September 1996 during Galileo's second fly-by of Ganymede. The coordinate system is defined in Fig. 2 legend. A short gap in the data removes artefacts associated with a programmed gain change. (The gap is covered in the recorded data.) The coordinate system is the same as in Fig. 2. Galileo's radial distance from the centre of Ganymede varied from $3.67 R_G$ at 18:40 UT to $\sim 1.10 R_G$ at closest approach and out to $3.74 R_G$ at 19:20 UT. Dashed lines show the superposition of the model fields discussed in the text. The data resolution is ~ 1 minute.



The source of Ganymede's magnetic field could be dynamo action in a molten iron core (or a salty-water internal ocean) or remanent magnetization in its interior. Ganymede's internal structure and thermal state determine which of these possibilities is the most likely. Schubert *et al.*¹² have considered how the field might be generated by dynamo action in an iron core or a watery mantle. The magnitude of Ganymede's magnetic field, almost an order of magnitude larger at the satellite's pole than the ambient jovian magnetic field, is in accord with theoretical expectations for a dynamo-generated field as is the direction of the dipole axis approximately aligned with the rotation axis¹². The dominantly

dipolar character of the field is consistent with generation deep within the body¹³. The large ratio between the polar field and the ambient field of the jovian magnetosphere at Ganymede cannot be explained by simple reactive processes such as paramagnetism or magneto-convection.

Remanent magnetization, though unlikely, cannot be ruled out completely. Whether in a core below the Curie temperature or in a magnetized shell in the interior of the moon, iron-rich material with a magnetic moment per unit volume $\mu = 40\text{--}80 \text{ A m}^{-1}$ would produce the observed dipole moment. The external field required to account for such strong magnetization is in excess of the field

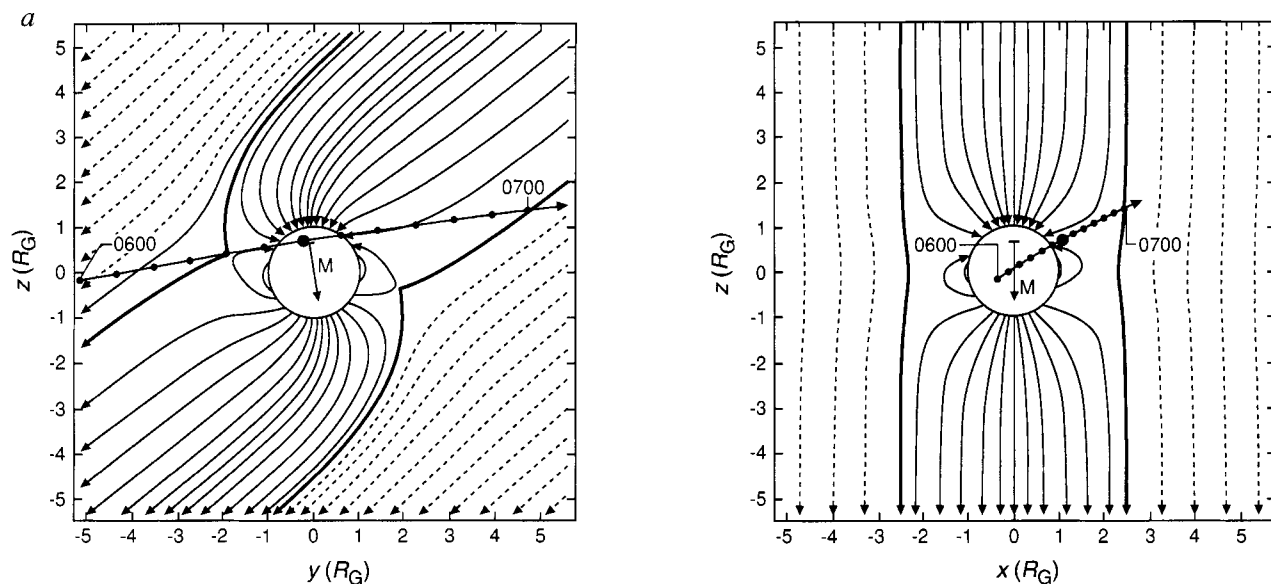


FIG. 5 Field lines in a vacuum superposition of a uniform external magnetic field of 120 nT lying in the y - z plane and tilted by 50° outward from $-z$ (a good approximation to the local value of the KK96 model field) and the field of a Ganymede-centred magnetic dipole with equatorial surface-field strength of 700 nT tilted 10° inward from $-z$. *a*, Field lines in the y - z plane. *b*, Field lines in the x - z plane. In both cases the projection of Galileo's trajectory is shown with dots every 5 minutes along the trajectory from 06:00 to 07:00 UT on 27 June 1996. Closest approach is marked with a

large dot. Note that the projection in *a* appears to place closest approach at high latitude, but the projection in *b* makes clear that the actual latitude is much lower. The separatrices between field lines linked to Ganymede and field lines that do not intersect the moon, effectively the magnetopause, are shown as heavy lines. Dashed lines are jovian field lines not linked to Ganymede. Solid lines are field lines with at least one end on Ganymede. An arrow marked M is aligned with Ganymede's magnetic moment.

near Ganymede in its present environment, but cannot be ruled out during its evolution. The inferred values cannot be rejected as impossibly large because the natural remanent magnetization of iron-rich meteorites can be this large in some cases.

We note that the presence of a strong internal field carries significant implications for the form of plasma flow near Ganymede. There should be no direct access of torus plasma to the surface other than in regions near the poles. Thus, deposits of sulphur-rich material from the torus should be localized near the poles. Sputtering of neutral atoms or molecules from Ganymede or its atmosphere would occur only in the polar regions, and this might inhibit significant neutral cloud formation and would certainly impose constraints on the cloud shape near the source. □

Received 30 September; accepted 5 November 1996.

1. Kivelson, M. G. *et al. Science* **273**, 337–340 (1996).
2. Kivelson, M. G., Khurana, K. K., Means, J. D., Russell, C. T. & Snare, R. C. *Space Sci. Rev.* **60**, 357–383 (1992).
3. Khurana, K. K. *J. Geophys. Res.* (submitted).
4. Connerney, J. E. P. in *Planetary Radio Emissions III* (eds Rucker, H. O., Bauer, S. J. & Kaiser, M. L.) 13–33 (Österreichischen Akademie der Wissenschaften, Vienna, 1992).
5. Neubauer, F. M. *J. Geophys. Res.* **85**, 1171–1178 (1980).
6. Southwood, D. J., Kivelson, M. G., Walker, R. J. & Slavin, J. A. *J. Geophys. Res.* **85**, 5959–5968 (1980).
7. Gurnett, D. A., Kurth, W. S., Roux, A., Bolton, S. J. & Kennel, C. F. *Nature* **384**, 535–537 (1996).
8. Hughes, W. in *Introduction to Space Physics* (eds Kivelson, M. G. & Russell, C. T.) 227–287 (Cambridge Univ. Press, New York, 1995).
9. Kivelson, M. G., Khurana, K. K., Walker, R. J., Wamecke, J. & Russell, C. T. *Science* **274**, 396–398 (1996).
10. Blackett, P. M. S. *Nature* **159**, 658 (1947).
11. Anderson, J. D., Shogren, W. L. & Schubert, G. *Science* **272**, 709–711 (1996).
12. Schubert, G., Zhang, K., Kivelson, M. G. & Anderson, J. D. *Nature* **384**, 544–545 (1996).
13. Elphic, R. C. & Russell, C. T. *Geophys. Res. Lett.* **5**, 211–214 (1978).
14. Dressler, A. J. in *Physics of the Jovian Magnetosphere* (ed. Dressler, A. J.) 498–504 (Cambridge Univ. Press, New York, 1983).

ACKNOWLEDGEMENTS. We thank S. Joy and A. Frederick for assistance in data preparation; R. L. Snare, J. Means, R. George, T. King and R. Silva for their varied contributions to the success of this effort; Y. Mei and D. Bindschadler for their support of magnetometer planning; V. Vasyliunas for discussions; and D. Stevenson for criticism. This work was supported by the Jet Propulsion Laboratory.

CORRESPONDENCE should be addressed to M.G.K. (e-mail: mkivelson@igpp.ucla.edu).

Gravitational constraints on the internal structure of Ganymede

J. D. Anderson*, E. L. Lau*, W. L. Sjogren*, G. Schubert† & W. B. Moore†

* Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109-8099, USA

† Department of Earth and Space Sciences, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90095-1567, USA

BEFORE the arrival of the Galileo spacecraft in the jovian system, there was little information on the interior structure of Jupiter's largest moon, Ganymede. Its mean density (1,940 kg m⁻³), determined by the Pioneer and Voyager spacecraft^{1–3}, implies a composition that is roughly 60% rock and 40% ice, which could be uniformly mixed or differentiated into a rocky core and icy mantle⁴. Here we report measurements by the Galileo spacecraft of Ganymede's overall density and the spherical harmonics, J_2 and C_{22} , of its gravitational field. These data show clearly that Ganymede has differentiated into a core and mantle. Combined with the recent discovery of an intrinsic magnetic field^{5,6}, our gravity results suggest that Ganymede has a metallic core of radius 400–1,300 km surrounded by a silicate mantle, which in turn enclosed by an ice shell ~800 km thick. Depending on whether the core is pure iron or an alloy of iron and iron sulphide, it could account for as little as 1.4% or as much as one-third of the

total mass. If the ice were stripped away, Ganymede could look much like Io⁷ in terms of its size and internal mass distribution.

The data were analysed by fitting a parametrized orbital model to the radio Doppler data by weighted nonlinear least squares^{8–10}. The two encounters between Galileo and Ganymede (on 27 June and 6 September 1996) were analysed independently. Ganymede's external gravitational field was modelled by the standard spherical harmonic representation of the gravitational potential¹¹. Because we assumed that the orientation of Ganymede's principal axes is known, only two non-zero coefficients (J_2 and C_{22}) were included in the model. All other harmonic coefficients were assumed to be exactly equal to zero. The two included coefficients measure the contributions to the gravitational potential of the spherical harmonics of degree l and order m for $l = 2, m = 0$ and $l = 2, m = 2$, respectively. In terms of spherical coordinates fixed in the body of Ganymede (radius r , latitude ϕ , and longitude λ), where longitude is measured from the Ganymede–Jupiter line in an equatorial system defined by Ganymede's spin axis, and r is the radial distance from Ganymede's centre of mass, the gravitational potential is

$$V = \frac{GM}{r} \left[1 - \frac{1}{2} J_2 \left(\frac{R}{r} \right)^2 (3 \sin^2 \phi - 1) + 3 C_{22} \left(\frac{R}{r} \right)^2 \cos^2 \phi \cos 2\lambda \right] \quad (1)$$

where G is the gravitational constant, and M and R are the mass and radius of the satellite. The two encounters were intentionally targeted to optimize gravitational field measurements; the first was a near-equatorial pass at an altitude of 835 km, while the second was a near-polar pass at an altitude of 261 km. The closest-approach location for the first was at $\phi = 30.394^\circ$ and $\lambda = 112.129^\circ$ (west longitude), while the second was at $\phi = 79.282^\circ$ and $\lambda = 122.444^\circ$. The first encounter was most sensitive to C_{22} , while the second was most sensitive to J_2 . However, for both encounters the two gravity coefficients were highly correlated, so we imposed the *a priori* hydrostatic constraint that J_2 is exactly 10/3 of C_{22} .

Gravity results for the two encounters are summarized in Table 1. Although the two gravity coefficients, because of the 10/3 constraint, were perfectly correlated *a priori*, the correlation after fitting the data is significantly less than unity for both encounters. This implies that there is considerable freedom for the two coefficients to differ from their hydrostatic values. To the contrary, it can be concluded from Table 1 that the non-hydrostatic perturbation to the second-degree field is only 0.2% with 1 σ uncertainty of 1.9%. It is highly unlikely that non-hydrostatic components contribute more than 6% to the Ganymede second-degree gravity field.

The value of GM is found to be $9,886.6 \pm 0.5 \text{ km}^3 \text{ s}^{-2}$. Masses derived from GM determinations depend on the gravitational constant G . The currently accepted value¹² is $(6.67259 \pm 0.00085) \times 10^{-20} \text{ km}^3 \text{ s}^{-2} \text{ kg}^{-1}$. This yields a mass of $(1.48167 \pm 0.00020) \times 10^{23} \text{ kg}$ for Ganymede. Under the assumption that the volume of Ganymede is equal to that of an equivalent sphere of radius $2,634 \pm 10 \text{ km}$ (ref. 13) its mean density is $1,936 \pm 22 \text{ kg m}^{-3}$, where the uncertainty is determined by the uncertainty in the radius.

Formal error estimates from the least-squares covariance matrix are based on an assumption of independent measurements drawn from a gaussian noise distribution. The reduced Galileo radio Doppler data are gaussian, but their power spectral density follows an $f^{-2/3}$ law arising from propagation of the radio carrier wave through solar plasma¹⁴. Here f is the Fourier frequency, not the 2.3 GHz frequency of the radiowave. The variance of spectral estimates of a signal roughly follows the same power-law dependence as the noise spectrum, so the Galileo gravity signals are better determined at higher Fourier frequencies. Our data weighting, with an assumed variance approximately equal to the variance of the Doppler residuals ($0.04 \text{ mm}^2 \text{ s}^{-2}$ at a sample interval of 60 s), is about right for the Ganymede gravity signals, where the peak

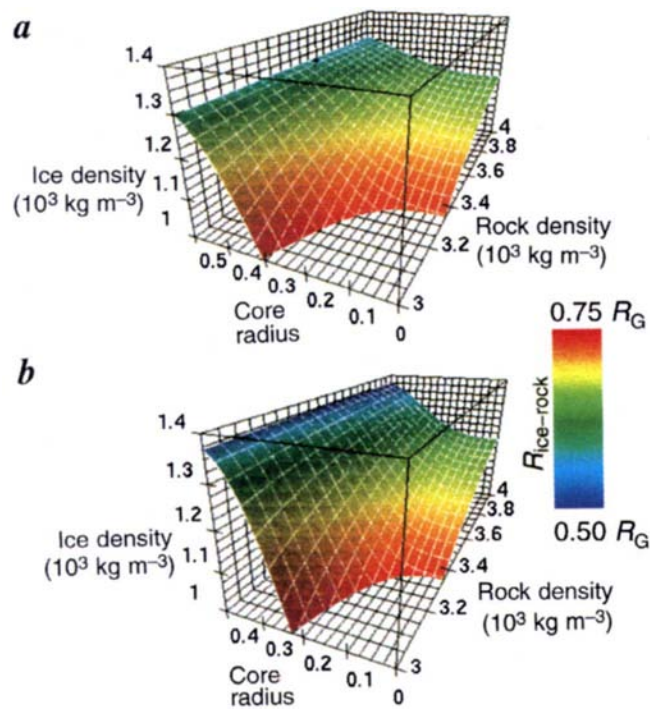


FIG. 1 Three-layer models of Ganymede consistent with observations of its overall density and C_{22} . *a*, Models with a core density of $5,150\text{ kg m}^{-3}$ (Fe-FeS); *b*, models with a core density of $8,000\text{ kg m}^{-3}$ (Fe). Any point on one of these surfaces defines a possible model of Ganymede's internal mass distribution with the values of ice density, rock density, and core-radius/Ganymede radius determined by the plots. The colours on the surfaces indicate the value of the ice-rock interface radius/Ganymede radius ($R_{\text{ice-rock}}/R_G$).

Fourier components are around $2 \times 10^{-3}\text{ Hz}$. We therefore retain the formal errors for the satellite gravity parameters for the first encounter. However, because of the closer approach of the spacecraft during the second encounter, we increase the formal errors by a factor of three in order to account for possible gravity perturbations by non-hydrostatic components. All errors reported here are our best estimates of realistic standard error.

Ganymede is in synchronous rotation with its orbital period¹³. If in addition it is in tidal and rotational equilibrium, it takes the shape of a triaxial ellipsoid with dimensions a, b and c ($a > b > c$). The long axis of the ellipsoid is along the planet-satellite line and the short axis is parallel to the rotation axis. The distortion of the satellite depends on the magnitude of the rotational and tidal forcing and the distribution of mass with radius inside the body. The distortion of the satellite and its internal mass distribution determine the satellite's gravitational field¹⁵⁻¹⁸. The gravitational coefficient C_{22} is related to the difference in the equatorial moments of inertia by

$$C_{22} = \frac{B - A}{4MR^2} \tag{2}$$

TABLE 1 Ganymede gravity results		
	Encounter 1	Encounter 2
J_2	$(126.0 \pm 6.0) \times 10^{-6}$	$(127.8 \pm 3.0) \times 10^{-6}$
C_{22}	$(37.8 \pm 1.8) \times 10^{-6}$	$(38.3 \pm 1.0) \times 10^{-6}$
μ	0.7399	0.5870

J_2 and C_{22} are defined in the text; μ is the correlation coefficient.

where the ellipsoidal satellite's principal moments of inertia are A, B and C ($C > B > A$). For a body in rotational and tidal equilibrium, the gravity coefficient C_{22} is related to the rotational parameter q_r by

$$C_{22} = \frac{3}{4} \alpha q_r \tag{3}$$

with similar equations for J_2 . Here α is a dimensionless response coefficient that depends on the distribution of mass within the satellite ($\alpha = 1/2$ for constant density) and q_r is the ratio of centrifugal to gravitational acceleration at the equator ($q_r = 1.903 \times 10^{-4}$ for Ganymede). For purposes of geophysical interpretation, we adopt the weighted mean of the two values of C_{22} in Table 1. The result is $C_{22} = (38.18 \pm 0.87) \times 10^{-6}$ and the corresponding value of α from equation (3) is 0.2675 ± 0.0061 (the value from the weighted mean of J_2 is $\alpha = 0.2679 \pm 0.0056$). The satellite's axial moment of inertia, normalized to MR^2 , follows from equilibrium theory. We obtain $C/MR^2 = 0.3105 \pm 0.0028$.

The axial moment of inertia provides a direct constraint on the internal mass distribution^{7,19}. For a uniform density body $C/MR^2 = 0.4$. The smaller the value of C/MR^2 , the more concentrated is the body's mass towards its centre. We note that the value of C/MR^2 for Ganymede is among the smallest of any planet or satellite in the Solar System. Io's value⁷ of C/MR^2 is 0.378; Earth's value is 0.334. Only the giant outer planets have C/MR^2 values smaller than Ganymede's value. Accordingly, it is immediately clear that Ganymede is strongly differentiated with a large concentration of mass toward its centre.

A more quantitative description of Ganymede's internal mass distribution can be given in terms of a model of the interior density $\rho(r)$. Consistent with the small number of constraints we can apply to the density distribution, the overall density and the value of C_{22} , we adopt a three-layer model with a core and two overlying spherical shells. A two-layer model is of course a special case of the three-layer model and corresponds either to a zero-thickness outer shell or a zero-radius core in the three-layer model. We solve Clairaut's equation²⁰ for the distortion of the model to the tidal and rotational potentials and determine the family of model parameters consistent with the observed value of α . There are more model parameters than available constraints, even for a two-layer model, and no unique model of Ganymede's internal mass distribution can be determined. Instead, we must restrict the model parameter space with reasonable assumptions about the nature of Ganymede's interior. We have emphasized above that Ganymede's moment of inertia requires that it is strongly differentiated and its overall density requires that it has a large water-ice

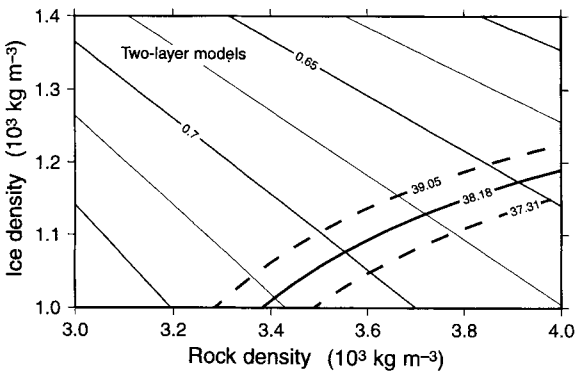


FIG. 2 Two-layer models of Ganymede consistent with its overall density and C_{22} . The curves labelled 39.05, 38.18 and 37.31 define possible Ganymede models for the nominal and plus and minus 1σ values of $C_{22} = (38.18 \pm 0.87) \times 10^{-6}$. The other curves give the value of ice-rock radius/Ganymede radius.

component. The existence of a magnetic field⁵ suggests that Ganymede has a metallic core. We therefore consider models in which the core density is either the density of Fe ($8,000 \text{ kg m}^{-3}$) or the density of Fe–FeS ($5,150 \text{ kg m}^{-3}$). We also assume that the outer shell of the model is predominantly water ice.

Allowable three-layer models consistent with Ganymede's overall density and C_{22} are shown in Fig. 1 for the two assumed values of core density. There are three additional model parameters that are unknown *a priori*, the densities of the ice and rock shells and the radius of the metallic core. The surfaces shown in the figure delineate the acceptable combinations of these parameters; the colours on the surface give the radius of the ice–rock interface. (Note that the radius of the core and of the ice–rock interface are given in units of R_G , the radius of Ganymede.) Two-layer models can be explored by the intersections of the surfaces with the plane representing core radius equal to zero. Because the core density is of no consequence when the core radius is zero, there is only one distinct intersection. The intersection is shown in Fig. 2 for the models based on the adopted value of C_{22} and for additional models not shown in Fig. 1 that correspond to the 1σ uncertainties in C_{22} . Two-layer models have rock densities larger than $\sim 3,400 \text{ kg m}^{-3}$ (in two-layer models, rock refers to silicates and metal), with the rock–ice interface at radii between about 0.64 and 0.73 of Ganymede's radius. Rock densities in excess of

$3,800 \text{ kg m}^{-3}$ are probably too large to be consistent with a silicate–metal mixture at the temperatures and pressures of Ganymede's interior, at least for approximately solar abundances of the relevant elements. However, rock densities between about 3,400 and $3,800 \text{ kg m}^{-3}$ are plausible. For example, if the metal and silicates in Io were rehomogenized and subject to ganymedean conditions, the density of the resultant mixture might be $3,600$ or $3,700 \text{ kg m}^{-3}$ (D. J. Stevenson, personal communication). Thus, while two-layer models of Ganymede's interior are possible, just on the basis of the gravity data, we favour instead a three-layer structure of Ganymede with a metallic core, in light of the existence of a ganymedean magnetic field^{5,6,21}.

For such a three-layer model of Ganymede, Fig. 3 shows the intersections of the surfaces in Fig. 1 with the plane representing a rock density of $3,300 \text{ kg m}^{-3}$, a reasonable density for the silicates alone. The figure also shows similar intersections for models based on the 1σ uncertainties in C_{22} . For the Fe–FeS core (core density, $5,150 \text{ kg m}^{-3}$), the core radius cannot exceed $\sim 0.5 R_G$ (that is, a fractional core radius of 0.5) without the complete disappearance of the silicate mantle. Possible models of Ganymede have cores with radii between about $0.2 R_G$ and $0.5 R_G$, masses between about 2% and 33% of Ganymede's mass, ice densities between about $1,000$ and $1,300 \text{ kg m}^{-3}$, and rock–ice interfaces at radii between about $0.6 R_G$ and $0.73 R_G$. For similar parameters and an Fe core, Fig. 3 indicates that the core radius is between about $0.15 R_G$ and $0.4 R_G$, the core mass is between about 1.4% and 26% of Ganymede's mass, the ice density is between about $1,000$ and $1,350 \text{ kg m}^{-3}$, and the rock–ice interface is between about $0.53 R_G$ and $0.73 R_G$. If the fractional core radius is larger than ~ 0.4 in these models, there is no silicate mantle.

The formation of a metallic core in Ganymede requires heating of the satellite to at least the Fe–FeS eutectic melting temperature ($\sim 1,325 \text{ K}$) at some time in the past. Accretional and radiogenic sources can provide this level of heating^{4,21}, but not much more. Ganymede could also have been tidally heated during passage through a temporary resonance in its orbital and thermal evolution^{21,22}. Io, which is in a resonance with Ganymede and Europa, is at present tidally heated^{14,23,24} and it may have been differentiated in the past by this mechanism. Io and Ganymede may not only have structural similarities, but they may have experienced similar heating and differentiation episodes in the past, even though Ganymede is not at present tidally heated. □

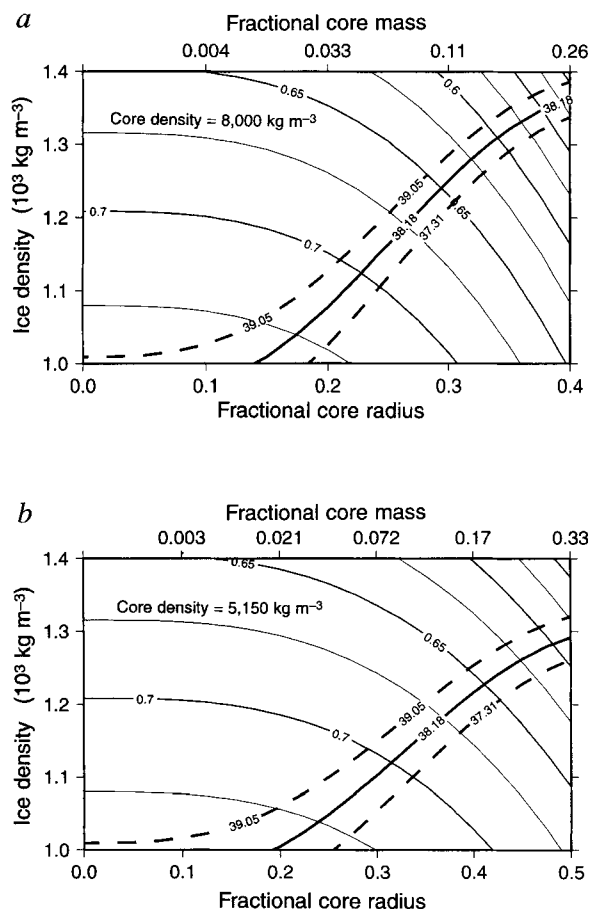


FIG. 3 Three-layer models of Ganymede with a rock density of $3,300 \text{ kg m}^{-3}$. *a*, Assuming an iron core with density $8,000 \text{ kg m}^{-3}$; *b*, assuming a Fe–FeS core with density $5,150 \text{ kg m}^{-3}$. The curves labelled 39.05, 38.18 and 37.31 define possible Ganymede models for the nominal and plus and minus 1σ values of $C_{22} = (38.18 \pm 0.87) \times 10^{-6}$. The other curves give the value of ice–rock radius/Ganymede radius. The upper horizontal scale is a nonlinear scale giving the fractional core mass (core mass/Ganymede mass). The fractional core radius is core-radius/Ganymede radius.

Received 8 October; accepted 12 November 1996.

1. Null, G. W., Anderson, J. D. & Wong, S. K. *Science* **12**, 476–477 (1975).
2. Null, G. W. *Astron. J.* **81**, 1153–1161 (1976).
3. Campbell, J. K. & Synnott, S. P. *Astron. J.* **90**, 364–372 (1985).
4. Schubert, G., Spohn, T. & Reynolds, R. T. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 629–688 (Univ. Arizona Press, Tucson, 1986).
5. Kivelson, M. G. et al. *Nature* **384**, 537–541 (1996).
6. Gurnett, D. A., Kurth, W. S., Roux, A., Bolton, S. J. & Kennel, C. F. *Nature* **384**, 535–537 (1996).
7. Schubert, G., Spohn, T. & Reynolds, R. T. *Science* **272**, 709–712 (1996).
8. Moyer, T. D. Tech. Rep. No. RT32-1527 (Jet Propulsion Lab., Pasadena, 1971).
9. Anderson, J. D. in *Experimental Gravitation* (ed. Bertotti, B.) 163–1299 (Academic, New York, 1974).
10. Tapley, B. D. in *Recent Advances in Dynamical Astronomy* (eds Tapley, B. D. & Szebehely, V.) 396–425 (Reidel, Dordrecht, 1973).
11. Kaula, W. M. *Theory of Satellite Geodesy* (Blaisdell, Waltham, MA, 1966).
12. Cohen, E. R. & Taylor, B. N. *Phys. Today* **49**, BG9–BG13 (1996).
13. Davies, M. E. et al. *Celest. Mech. Dyn. Astron.* **53**, 377–397 (1992).
14. Woo, R. & Armstrong, J. W. *J. Geophys. Res.* **84**, 7288–7296 (1979).
15. Kaula, W. M. *An Introduction to Planetary Physics: The Terrestrial Planets* (Wiley, New York, 1968).
16. Hubbard, W. B. & Anderson, J. D. *Icarus* **33**, 336–341 (1978).
17. Dermott, S. F. *Icarus* **37**, 310–321 (1979).
18. Zharkov, V. N., Leontjev, V. V. & Kozenko, A. V. *Icarus* **61**, 92–100 (1985).
19. Schubert, G., Limonadi, D., Anderson, J. D., Campbell, J. K. & Giampieri, G. *Icarus* **111**, 433–440 (1994).
20. Jeffreys, H. *The Earth* (Cambridge Univ. Press, 1962).
21. Schubert, G., Zhang, K., Kivelson, M. G. & Anderson, J. D. *Nature* **384**, 544–545 (1996).
22. Malhotra, R. *Icarus* **94**, 399–412 (1991).
23. Segatz, M., Spohn, T., Ross, M. N. & Schubert, G. *Icarus* **75**, 187–206 (1988).
24. Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892–894 (1979).

ACKNOWLEDGEMENTS. This work was supported by the Galileo Project and was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. G.S. and W.B.M. were supported by NASA through the Galileo Project at JPL and the Planetary Geology and Geophysics Programs.

CORRESPONDENCE should be addressed to J.D.A. (e-mail: john.d.anderson@jpl.nasa.gov).

The magnetic field and internal structure of Ganymede

Gerald Schubert^{*†}, Keke Zhang[‡],
Margaret G. Kivelson^{†*} & John D. Anderson[§]

^{*} Department of Earth and Space Sciences, [†] Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90095, USA

[‡] Department of Mathematics, University of Exeter, Exeter EX4 4QJ, UK

[§] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

BEFORE the recent fly-bys of Ganymede by the Galileo spacecraft, viable models of the internal structure of Jupiter's largest moon ranged from a uniform mixture of rock and ice to a differentiated body with a rocky core and an icy mantle¹. Analysis of the Doppler shift of radio signals from Galileo has now shown that Ganymede is strongly differentiated with a relatively dense core surrounded by a thick shell of ice². Other instruments have revealed that Ganymede has an intrinsic magnetic field^{3,4}, which is aligned approximately antiparallel to Jupiter's magnetic field. Here we argue that these results imply that Ganymede has an outer silicate core surrounding a liquid (or partially liquid) inner core of iron or iron sulphide, and that the magnetic field is generated by dynamo action within this metallic core. Io also appears to have an intrinsic magnetic field^{5,6} (antiparallel to that of Jupiter), implying that it too has a metallic core⁷ in which the field is generated either by dynamo action or by magneto-convection.

The magnetic fields of Ganymede and Io could be due to either remanent magnetization, dynamo action, or magneto-convection in the satellites' interiors. Which is the more likely source of the magnetic field of each satellite depends on the structures and thermal states of the satellites. We consider first the possibility of remanent magnetization in both satellites. Because a magnetized layer must have temperatures below the Curie temperature, the most favourable location in Ganymede for such a layer would be at the top of a rock (metal silicate) core just below the ice mantle. The heat flux at the rock-ice interface in Ganymede is estimated to be $\sim 10 \text{ mW m}^{-2}$, assuming that the interface has a radius of $\sim 70\%$ of Ganymede's radius and a silicate heat source density⁸ of $1.8 \times 10^{-8} \text{ W m}^{-3}$. The temperature at the ice-rock interface in Ganymede is probably close to the ice melting point, leaving about 600–700 K for the additional temperature rise to the Curie point. With the assumption that the silicate thermal conductivity in Ganymede is $\sim 4 \text{ W m}^{-1} \text{ K}^{-1}$, typical of rocks in the Earth's mantle, we find that the maximum thickness of a magnetized rock shell just below the ice-rock interface is $\sim 300 \text{ km}$. Although it is possible to achieve the requisite level of magnetization in such a layer to account for Ganymede's intrinsic magnetic dipole moment^{3,9}—for example, by a sufficiently large concentration of magnetite and an external magnetizing field larger than the present jovian magnetic field at Ganymede's orbit—the requirements fall outside reasonable parameter values.

The possibility that Io's magnetic field is due to remanent magnetization seems unlikely because Io is intensely heated by tidal dissipation^{1,10,11} and any magnetized layer in its mantle would have temperatures below the Curie temperature. Io's large surface heat flux¹² ($\sim 2.5 \text{ W m}^{-2}$) limits the thickness of an outer magnetized shell to $\sim 2 \text{ km}$, assuming a temperature rise of $\sim 1,000 \text{ K}$ between the temperature of the surface and the Curie point and a thermal conductivity of $4 \text{ W m}^{-1} \text{ K}^{-1}$. Even for a shell 100 km thick, the magnetization would have to be unreasonably large (or the amount of free iron in the shell would have to be excessively large) to account for the magnetic dipole moment of Io inferred from the Galileo observations⁵. Even so, Cheng and Paranicus¹³ have proposed that Io's magnetic field is due to

remanent magnetization. They weaken the above argument against remanent magnetism by hypothesizing that the thermal conductivity in the outer several hundred kilometres of Io is substantially larger than that of typical silicates because of the pervasive presence of sulphur-rich magmas near Io's surface. These magmas supply the surface volcanics and transport Io's internal heat. However, the existence of these magmas and other evidence for a partially molten asthenosphere¹⁴ near Io's surface imply temperatures well in excess of the Curie temperature at depths of $\sim 100 \text{ km}$.

Although remanent magnetization cannot be ruled out with certainty as the source of either Ganymede's or Io's intrinsic magnetic field, we consider it more likely that the magnetic fields of these satellites are generated by dynamo action or magneto-convection in internal electrically conducting liquid regions. Dynamo action is the magnetohydrodynamic process that converts mechanical energy of convective motions in an electrically conducting fluid into magnetic energy. A dynamo maintains a magnetic field against ohmic dissipation; the self-sustaining magnetic field of a dynamo exists so long as the drive for convection (thermal or compositional) continues. A dynamo does not require an external magnetic field in order to function, but it may be influenced by such a field¹⁵. A dynamo requires the magnetic Reynolds number Re_m (given by $Re_m = UL/\eta$, where U is a representative convective velocity, L is a representative length scale and η is the magnetic diffusivity) to exceed a critical value that has been shown to be greater than about 10. The amplitude of the generated magnetic field is determined by the form and magnitude of convection¹⁶.

Magneto-convection refers to convective motions of an electrically conducting fluid that are affected by an imposed magnetic field. The motions and associated currents change and perturb the imposed magnetic field. The perturbation magnetic field is of the same order of magnitude as the imposed field, so the new total magnetic field is also generally about the size of the imposed field but in places it could be much smaller. In magneto-convection there is no self-sustenance of the magnetic field, that is, if the external magnetic field were removed, convection would not by itself generate a new magnetic field. There is no critical value for the magnetic Reynolds number, which may be of the order of unity or smaller for magneto-convection.

In the case of Ganymede there are two possible regions of electrically conducting liquid, an ocean of salty water at the bottom of the satellite's ice shell or a metallic inner core. Models of Ganymede's internal temperature based on heat transfer by thermal conduction have an internal water ocean at the rock-ice interface¹⁷, but effects of convective heat transfer in the ice can preclude formation of such a layer or refreeze it, depending on the subsolidus rheology of the ice. If such an internal salt-water ocean existed, the condition $Re_m > 10$ for dynamo action would give $U > 1 \text{ m s}^{-1}$ for the electrical conductivity of salt water and a layer thickness of several hundred kilometres. By way of comparison, the convective motions in the Earth's core¹⁸ have a value of U of $\sim 10^{-4} \text{ m s}^{-1}$. The large velocities required for dynamo action in an internal ganymedeian salt water ocean seem improbable and are several orders of magnitude in excess of those predicted by Kolmogorov scaling (D. J. Stevenson, personal communication). Rejection of electrical currents in a hypothetical ganymedeian internal salty ocean as the source of the satellite's intrinsic magnetic field leaves dynamo action or magneto-convection in a liquid or partially liquid metallic core as the only reasonable source of Ganymede's magnetic field. We therefore conclude that Ganymede has an inner metallic core surrounded by an outer silicate core, a result that cannot be deduced from the gravity data² alone.

Dynamo action rather than magneto-convection is the process likely to be occurring in Ganymede's inner metallic core because Ganymede's magnetic field³ is so much larger (by a factor of six) than the ambient jovian magnetic field¹⁹ (the main jovian field and the field of the current sheet at Ganymede's orbit, averaged over a

Jupiter rotation). The dynamo process in Ganymede must be fundamentally different from the dynamo process in the Earth because of the presence of the ambient jovian magnetic field around Ganymede. For the Earth's dynamo, there is no background magnetic field, the dynamo equations are invariant under the transformation $\mathbf{B} \rightarrow -\mathbf{B}$ and there is no preferred direction to the self-generated magnetic field. For Ganymede's dynamo, the magnetohydrodynamic equations contain additional terms due to the jovian ambient magnetic field and transformation invariance is no longer valid. The ambient jovian magnetic field provides a preferred direction for the ganymede dynamo as does Ganymede's spin vector. These directions approximately coincide because the appropriate time-averaged background magnetic field is nearly in the direction of Ganymede's rotation vector (R. Holme, personal communication).

The condition $Re_m > 10$ for dynamo action in an inner metallic ganymede core gives $U > 2 \times 10^{-5} \text{ m s}^{-1}$ for a length scale of $\sim 500 \text{ km}$ (a possible value for the inner core radius) and a magnetic diffusivity $\eta = 1 \text{ m}^2 \text{ s}^{-1}$ (a typical value for the Earth's core). Because the electrical conductivity of molten iron is more than four orders of magnitude larger than that of salt water, the convective velocities in a liquid metallic inner core are smaller than those required in a salty ocean by the same factor. The estimated velocities in the inner metallic core are reasonable when compared to the estimates of similar velocities in the Earth's core.

In the case of Io, the obvious location for dynamo action or magneto-convection is the satellite's metallic core⁷ though the asthenosphere¹⁴ is another possibility assuming it is sufficiently molten. If dynamo action were occurring in Io's core, then the above estimate of convective velocities in Ganymede's inner core would also apply to Io as the cores of these satellites could have similar dimensions. In fact, a model of Io surrounded by $\sim 800 \text{ km}$ of ice would satisfy the gravity data for Ganymede². As the magnetic dipole moments of Io and Ganymede are about the same^{3,5} it would be reasonable to conclude that both moons have similarly energetic dynamos. However, there are a number of reasons to suspect that Io's core may not have an active dynamo but instead be magneto-convective. The magnitude of Io's inferred magnetic dipole moment⁵ is actually an upper bound because current systems in the plasma surrounding Io and induction due to the passage of Io through Jupiter's magnetic field make an as-yet undetermined contribution to the observed magnetic field^{5,6,20}. Thus Io's magnetic dipole moment could be smaller than Ganymede's dipole moment. The ionian magnetic field at Io's surface is about the same magnitude (or smaller if plasma currents are important) as the jovian background field. Although this could be fortuitous, it is consistent with magneto-convection rather than dynamo action in Io's core. Theoretical considerations²¹ point to a difficulty in sustaining an active dynamo in Io's core if its mantle is at present being intensely heated by tidal dissipation^{1,10,11}. Tidal heating in Io's mantle retards the cooling and solidification of the core and the thermal and compositional core convection that could sustain a dynamo²¹. An ionian dynamo would be expected only during times in Io's evolution when strong tidal heating of the mantle was not occurring²¹. Accordingly, we consider it plausible that Io's core is in a magneto-convective state or, if dynamo action is taking place, that the dynamo is a weak one, that is, its magnetic Reynolds number only slightly exceeds the critical value for the onset of dynamo action. The strong tidal heating of the mantle that retards core cooling and dynamo action in Io does not at present take place in Ganymede.

The ambient jovian magnetic field must play an essential role in magneto-convection in Io's core. The effect is to relax the rotational constraint on the dynamics and thus to enhance convection. To achieve an equilibrium it would thus be expected that the amplitude of the imposed magnetic field would be reduced by the convective modification in agreement with the magnetic-field observations⁵ near Io. The influence of the ambient

jovian magnetic field on magneto-convection in Io and dynamo action in Ganymede provides a new direction of enquiry that may elucidate the general nature of dynamo behaviour. It is possible, for example, that as a consequence of the ambient jovian magnetic field, the ganymede dynamo cannot undergo field reversals.

For the magnetic fields of Io and Ganymede to originate in the metallic cores of the satellites, it is necessary that the cores be molten, or at least partly liquid. This constrains the temperatures at the iron-core/silicate boundaries of the satellites to be at or above the melting temperature of the core materials. For a core of pure iron, the melting temperature²² at the pressure of the core-silicate boundary is about 1,920 K for Io and 1,980 K for Ganymede; for a core of Fe-FeS these temperatures²³ are about 1,250 K and 1,325 K, for Io and Ganymede, respectively. These temperatures place important constraints on the entire thermal structures of the satellites and provide end points for their thermal histories.

In the case of Io, the existence of a core was expected¹¹, but the finding of a magnetic field was generally not anticipated. In the case of Ganymede, both the magnetic field and metallic core come as surprises. For Io, accretional, radiogenic and especially tidal heating¹⁰ provide a surfeit of energy for differentiation of a metallic core, but for Ganymede, accretional heating is probably insufficient for early formation of a metallic core though it may be adequate to separate the ice from the rock and metal¹. Radiogenic heating during the course of Ganymede's evolution could provide enough heating to separate an Fe-FeS core if the eutectic melting temperature of about 1,250 K is all that is required²⁴. Although formation of a metallic core in Ganymede may not be energetically difficult (but we note that it was not anticipated before the Galileo fly-by), there may be a more serious problem involved in explaining how the core has remained convective throughout Ganymede's history as small cores tend to evolve toward an isothermal state (D. J. Stevenson, personal communication). Although tidal heating is unimportant in Ganymede at present, tidal dissipation during passage through a dynamical resonance at some earlier time in its evolution²⁵ may have played a role in sustaining or restarting core convection or even in core formation or maintenance of the liquid state. Detailed dynamical-thermal evolution models of Ganymede are needed to study these possibilities. $\square$

Received 7 October; accepted 8 November 1996.

1. Schubert, G., Spohn, T. & Reynolds, R. T. In *Satellites* (eds Burns, J. A. & Matthews, M. S.) 224–292 (University Arizona Press, Tucson, 1986).
2. Anderson, J. D., Lau, E. L., Sjogren, W. L., Schubert, G. & Moore, W. B. *Nature* **384**, 541–543 (1996).
3. Kivelson, M. G. et al. *Nature* **384**, 537–541 (1996).
4. Gurnett, D. A., Kurth, W. S., Roux, A., Bolton, S. J. & Kennel, C. F. *Nature* **384**, 535–537 (1996).
5. Kivelson, M. G. et al. *Science* **273**, 337–340 (1996).
6. Kivelson, M. G. et al. *Science* **274**, 396–398 (1996).
7. Anderson, J. D., Sjogren, W. L. & Schubert, G. *Science* **272**, 709–712 (1996).
8. Turcotte, D. L. & Schubert, G. *Geodynamics* (Wiley & Sons, New York, 1982).
9. Cray, F. J. & Bagenal, F. (abstr.) *Bull. Am. Astron. Soc.* **28**, 1075 (1996).
10. Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203**, 892–894 (1979).
11. Segatz, M., Spohn, T., Ross, M. N. & Schubert, G. *Icarus* **75**, 187–206 (1988).
12. Veeder, G. J., Matson, D. L., Johnson, T. V., Blaney, D. L. & Gougen, J. D. *J. Geophys. Res.* **99**, 17095–17162 (1994).
13. Cheng, A. F. & Paranicus, C. *Geophys. Res. Lett.* **23**, 2879–2882 (1996).
14. Ross, M. N., Schubert, G., Spohn, T. & Gaskell, R. W. *Icarus* **85**, 309–325 (1990).
15. Levy, E. H. *Proc. Lunar Planet. Sci. Conf.* **10**, 2335–2342 (1979).
16. Stevenson, D. J. *Rep. Prog. Phys.* **46**, 555–620 (1983).
17. Moore, W. B. & Schubert, G. (abstr.) *Bull. Am. Astron. Soc.* **28**, 1139 (1996).
18. Bloxham, J. & Jackson, A. *Rev. Geophys.* **29**, 97–120 (1991).
19. Khurana, K. K. *J. Geophys. Res.* (in the press).
20. Frank, L. A. et al. *Science* **274**, 394–395 (1996).
21. Wienbruch, V. & Spohn, T. *Planet. Space Sci.* **43**, 1045–1057 (1995).
22. Sterrett, K. F., Klement, W. Jr & Kennedy, G. C. *J. Geophys. Res.* **70**, 1979–1984 (1965).
23. Usselman, T. M. *Am. J. Sci.* **275**, 278–290 (1975).
24. McKinnon, W. B. (abstr.) *Bull. Am. Astron. Soc.* **28**, 1076 (1996).
25. Malhotra, R. *Icarus* **94**, 399–412 (1991).

ACKNOWLEDGEMENTS. We thank W. B. Moore for assistance in carrying out the thermal models and D. J. Stevenson and R. Holme for helpful reviews. This work was supported by NASA through the Galileo Project at JPL and through the Planetary Geology and Geophysics Program.

CORRESPONDENCE should be addressed to G.S. at the Department of Earth and Space Sciences (e-mail: schubert@ucla.edu).

Relationship between melting and amorphization of ice

Osamu Mishima

National Institute for Research in Inorganic Materials, 1-1, Namiki, Tsukuba, Ibaraki 305, Japan

THE discovery¹ in 1984 that an ice crystal can be transformed by pressure to an amorphous phase has since been followed by other examples of pressure-induced amorphization². This transition, like melting, involves loss of long-ranged order, prompting the question of whether the two transitions are related. Here I describe experiments probing this relationship for a form of crystalline ice (denoted Ih) which is melted and amorphized by pressure. To avoid the complication of crystal-crystal transformations interrupting the melting process I use an ice emulsion, in which the very small particle size (about 5 μm) suppresses nucleation of other crystal phases. As the temperature is decreased, I see a smooth crossover from (pressure-induced) equilibrium melting to sluggish amorphization at around 140–165 K. In this temperature range, ice Ih became ‘supercompressed’ before melting to a highly viscous liquid which seemed to be related to an imperfectly relaxed amorphous ice. Below about 140 K, ice Ih was transformed to an unrelaxed phase apparently related to the high-density amorphous form of ice. This sequence of transitions can be viewed as a crossover from a two-phase melting process (which is determined by the relative free energies of the solid and liquid phases) towards a one-phase amorphization process (where the transition is induced by a mechanical instability limit of the solid).

Between 273 and 250 K, ice Ih can be melted by applying pressure. When compressed at 77 K, ice Ih is amorphized¹. But the relationship between these two transitions is hard to probe because, below about 250 K, ice Ih is transformed under pressure to other crystalline forms (such as ice III and IX) before melting occurs. Molecular-dynamics calculations^{3,4} have provided some information on the relation between melting and amorphization, but experimental data are almost non-existent.

If the formation of these other crystal phases can be suppressed, however, ice Ih melts to metastable (supercooled) water below ~ 250 K. By conducting pressurization experiments on ice Ih in an emulsion (see Fig. 1 legend), I have mapped out the melting and amorphization curves between 250 and 77 K (Figs 1, 2). This can be divided into four regimes.

In region A ($250\text{ K} > T > 190\text{ K}$), ice Ih undergoes thermodynamic (metastable) melting to supercooled water (*in situ* X-ray characterization of the liquid phase would be desirable, but has not yet been obtained). The transition pressures extrapolate smoothly to the bulk melting line above ~ 250 K. The transition is endothermic (Fig. 3), and the transformed phase (supercooled emulsified water) could be recovered by decompressing to atmospheric pressure above 235 K. As the ice is compressed (quasi) adiabatically through the transition, the endothermic nature means that heat is absorbed by the melted fraction, the sample temperature decreases, and a higher pressure is needed to allow the transition to go to completion along the melting line of ice Ih. This shows that the melting is an equilibrium transition. I also observed melting of emulsified ice III, V and VI along their extrapolated melting lines, showing that this ability of emulsions to reveal metastable melting is quite general.

The crossover from region A to region B corresponds to the point at which other crystal phases can nucleate homogeneously in the supercooled water emulsion formed by melting of ice Ih. This homogeneous nucleation temperature of ~ 190 K at 0.5 GPa was estimated by extrapolating that measured between 0 and 0.3 GPa (ref. 5). So in region B, the ice Ih emulsions were observed to transform to ice V between 190 and 180 K, and to ice IX or III

below ~ 180 K. The latter occurred in two steps as the pressure was increased: gradually at first (a in Fig. 1, hatched area in Fig. 2a) and then quickly (b in Fig. 1, circles in Fig. 2a). The comparable sharpness of curve b in region B with melting at 197 K in region A (Fig. 1) suggests that the transitions in both regions A and B begin by the same mechanism, that is, melting of ice Ih. Moreover, the transitions at b extrapolate smoothly to the melting line in region A, and so it seems reasonable to regard them as the locus of a melting transition of ice Ih to supercooled water followed by freezing to ice V or IX. The transitions in region B are exothermic, consistent with the evolution of heat due to the net crystal-crystal (ice Ih to ice V or IX) transformation.

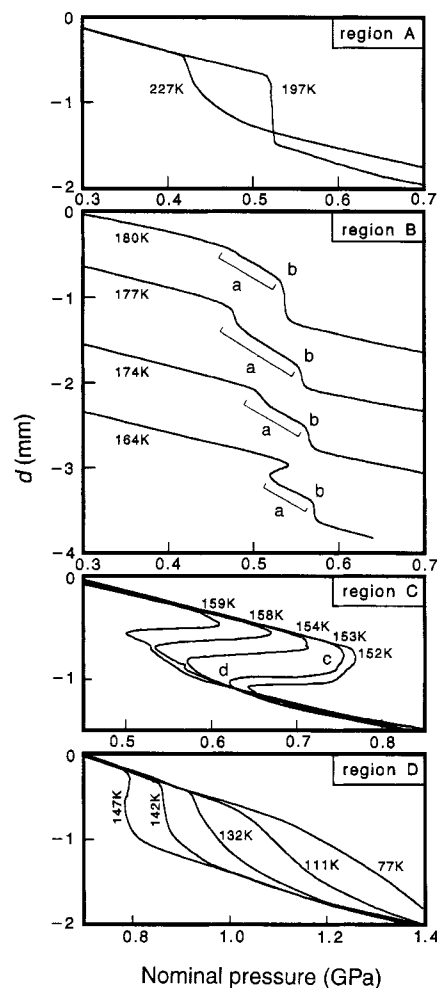


FIG. 1 Compression of fresh ice Ih emulsions (in terms of piston displacement d ; see below) as a function of nominal pressure for first-compression runs. The temperature around the onset of transition is indicated. Volume of ice Ih decreased by $\sim 20\%$ at the transitions in regions B, C and D, which agreed approximately with the change to ice V, IX or HDA. Compression of the emulsion carrier material between 77–273 K showed a monotonous decrease in volume.

METHODS. About 1.3 g of water emulsion⁵ (droplets 1–10 μm in size), made by stirring methylcyclohexane (3 cm^3), methylcyclopentane (3 cm^3), distilled water (4 cm^3) and sorbitan tristerate (200 mg) with a homogenizer, was loaded into an indium cup and compressed in a steel cylinder (bore diameter, 15 mm) by a hydraulic press⁹. The apparatus, cooled in a pail filled with liquid nitrogen, warmed gradually ($\sim 1.3\text{ K min}^{-1}$ at 100–150 K, $\sim 0.9\text{ K min}^{-1}$ at 150–200 K) after the nitrogen evaporated. During the heating, the sample was compressed up to ~ 1.5 GPa at a constant rate of 0.1 GPa min^{-1} (except at transitions). The piston displacement (d), the press load and cylinder temperature were measured during the experiment. The given temperatures at high pressure are the cylinder temperatures and have an error of ± 3 K. The pressure of the sample is $\sim 90\%$ of the nominal pressure.

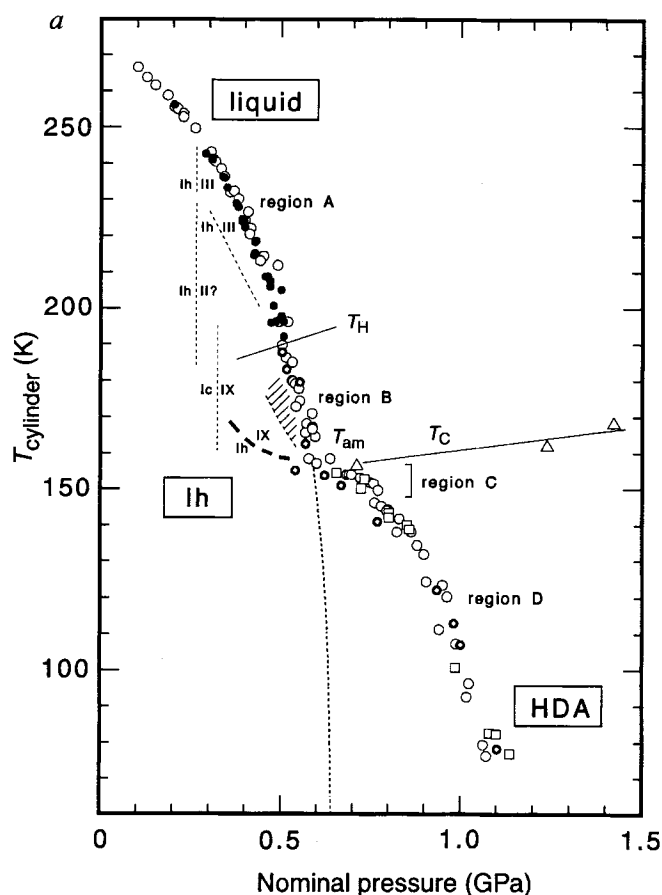
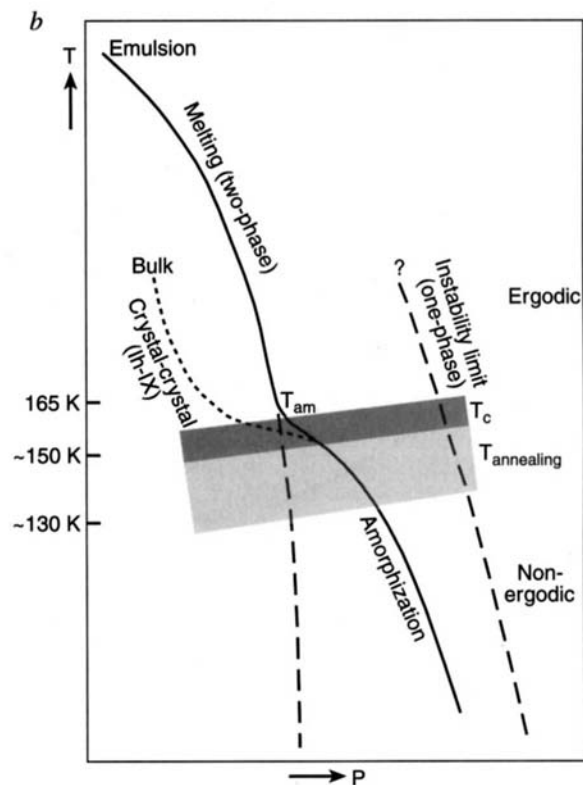


FIG. 2 a, The pressure-induced melting-amorphization line of ice Ih. Transition pressures were obtained by detecting a decrease in sample volume (empty circles, Ih emulsion; empty squares, pure bulk Ih) or a decrease (filled circles) or an increase (two concentric circles) in the sample temperature. When compression and decompression were repeated, crystal-crystal transitions (thin dotted lines) also occurred. They were distinguished from the melting transitions by their transition pressures, except near 245 K, 0.27 GPa and 200 K, 0.5 GPa. Around 200 K and 0.5 GPa, the Ih-III transition was exothermic, and the

In regions C ($\sim 160 \text{ K} > T > \sim 150 \text{ K}$) and D (below $\sim 150 \text{ K}$), I obtained the same results both for ice emulsions and for bulk ice Ih. In region C, I observed amorphization to high-density amorphous (HDA) ice followed by crystallization to ice IX. Above $\sim 160 \text{ K}$, bulk ice Ih is transformed to IX directly before melting (bold dashed line in Fig. 2a); so we can consider that in region C the HDA ice appears as an intermediate phase in this crystal-crystal transformation. The samples underwent 'over-pressurization' (Fig. 1, region C)—a transformation beginning at point c lowered the pressure to point d. The emulsion sample recovered at 1 bar from point c showed two exothermic transitions on heating—first to low-density amorphous (LDA) ice and then to crystalline ice Ic^{1,6}. The location of the first transition is consistent with its identification with the HDA $\rightarrow$ LDA transition, showing that HDA ice is present in the sample at c—although the X-ray diffraction pattern showed only untransformed ice Ih, the HDA halo being presumably too weak to detect. The sample from d showed the X-ray pattern of ice IX, and I suppose that the exothermic amorphization at c induces this crystallization.

Although the slope of the melting/amorphization curve changes rather abruptly from region B to C, I believe that the crossover here is continuous. Primarily this is because the shape of the compression curve of melting at b in region B below about 165 K is the same as that of amorphization in region C, with a slight over-



endothemic nature of the melting process revealed the onset melting pressure. The thick dotted line is the extrapolated thermodynamic melting line. The crystallization temperature of HDA (triangles and the T_c line) and the extrapolated homogeneous freezing temperature of supercooled water (the T_H line), are also shown. The hatched area corresponds to the gradual transition a in Fig. 1. The thick dashed line shows the ice Ih-IX transitions of pure bulk ice Ih. The melting line and the amorphization line merge around T_{am} . b, Schematic representation of the crossover from two-phase towards one-phase disordering.

pressurization and subsequent pressure relaxation. I therefore suggest that the melting transition in region B changes smoothly to amorphization (followed by crystallization) at a temperature T_{am} between 165 and 155 K in region C (Fig. 2b).

In region D, amorphization of ice Ih to HDA occurred without any subsequent crystallization—all recovered samples were identified as HDA by X-ray and thermal analysis. Thus, ice Ih melts in region A, melts and freezes in region B, amorphizes and crystallizes in region C and amorphizes in region D, with a continuous crossover between these behaviours.

Although all samples recovered from region D were HDA ice, their X-ray patterns and HDA-LDA transition temperatures varies slightly with amorphization temperature (Fig. 4). HDA ice formed by amorphization at 77 K and heated to $\sim 140 \text{ K}$ at 1.5 GPa resembled that obtained by amorphization at 140 K. Moreover, the HDA ices made by transformation not only of ice Ih¹ but also ice Ic⁷ and LDA ice^{8,9} under various pressure-temperature conditions all show identical X-ray patterns (with a halo peak around $2.75 \text{ \AA}$) and similar HDA-LDA transition temperatures (around 123–128 K) when heated or annealed to 130–150 K at 1.0–1.5 GPa. This shows that the different HDA ices are all in the same HDA 'megabasin' on the potential-energy surface¹⁰—that is, they are different metastable states that can be annealed to the same (relaxed) HDA potential minimum.

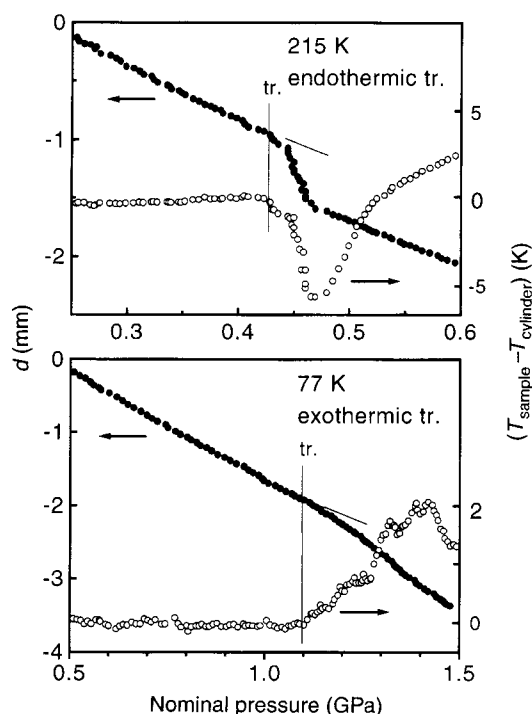


FIG. 3 Piston displacement d , and difference in temperature between the sample and the cylinder, during compression of ice Ih emulsions. Data were obtained at 215 K (top panel) and 77 K (bottom panel). The sample temperature was measured during compression (and during heating at a constant nominal pressure) by a thermocouple brought into the sample through a small hole in a piston⁹. When steadily heated, the sample temperature was lower than the cylinder temperature by ~ 4 K at ~ 140 K, ~ 2.5 K at ~ 160 K, and ~ 1 K at ~ 200 K. The transition (tr.) pressure was defined as the pressure at which the volume started to decrease where the sample temperature clearly began to change.

I performed differential thermal analysis (DTA) of preheated or annealed pure HDA ice at high pressures to determine the glass transition temperature T_g of the relaxed HDA ice—that is, the temperature at which the molecular motion of supercooled water ‘freezes out’ to form the glassy HDA phase. I observed pressure-dependent crystallization of HDA ice (an exothermic transition) at temperatures shown as T_c in Fig. 2, and the line of transition temperatures apparently crosses the melting/amorphization curve around T_{am} , which is consistent with the idea that amorphization occurs only below T_{am} , and melting followed by crystallization above it. But the DTA experiments showed no signature of an endothermic glass transition (to within $\pm \sim 0.02$ K) before crystallization, implying that either T_g is very close to T_c or the difference in heat capacity between HDA and supercooled water is too small to be detected. Nevertheless, the glass transition temperature should be above, or near, the temperature at which HDA ice can be annealed from metastable forms (that is, 130–150 K) and is usually below T_c (155–160 K), so I propose that it lies in the region 150–160 K at 0.6–0.8 GPa. This value is close to the ideal glass transition temperature derived for supercooled water at high pressure by fitting diffusion data to the Vogel–Tammann–Fulcher equation¹¹.

As the temperature decreases below T_{am} , the amorphization line bends such that amorphization goes from being more temperature-sensitive (between 160 and 140 K) to more pressure-sensitive (below ~ 140 K). This suggests that at higher pressures the potential barrier for amorphization decreases and some kind of mechanical instability of ice Ih is being approached⁴. In other words, the amorphization mechanism may change gradually from being a kind of kinetically delayed ‘melting’ (a two-phase picture) towards a collapse due to lattice instability (a one-phase picture). Thus, in region C we might suppose that the motion of water

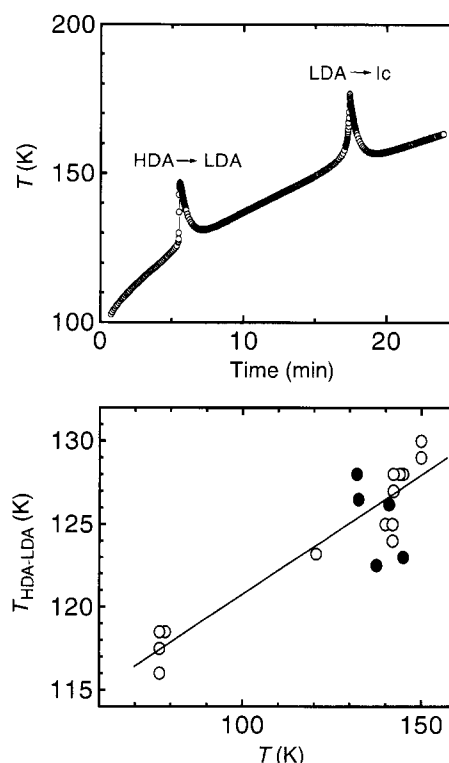


FIG. 4 Top panel, a typical heating curve of high-density amorphous (HDA) ice at 1 bar (LDA, low-density amorphous ice; Ic, ice Ic). Bottom panel, HDA–LDA transition temperatures of HDA ices. Open circles, HDA formed by amorphization at T K; filled circles, HDA heated (or annealed) to T K under high pressure. Sample grains (~ 150 mg) were loaded in a thin aluminum cup (~ 15 mg) placed inside a steel pipe which was heated from 77 K to room temperature (at a rate of about $2\text{--}4$ K min^{-1} around 120–160 K). The temperature of the cup was measured with an attached fine thermocouple. The HDA showed two exothermic transitions^{1,6}, whereas ice IX showed an exothermic transition to ice Ic^{16,17} at ~ 155 K. The samples were also studied by X-ray diffraction; a glass vacuum flask⁹ having a capillary for X-ray diffraction was used. X-ray photographs of sample fragments (0.2–0.5 mm in size) in the flask filled with liquid nitrogen were taken using Zr-filtered Mo radiation and Polaroid films. Crystalline phases were determined by their ‘ring’ patterns, which occurred on a background ‘halo’ derived from the glass capillary, the liquid nitrogen and the emulsion carrier materials. HDA and LDA were differentiated by the position and width of their principle diffraction halo rings^{1,8,9}. The HDA formed by amorphization at 145 K had an intense halo ring corresponding to an interplanar spacing of ~ 2.75 Å; this peak was absent from the X-ray pattern of HDA formed by amorphization at 77 K.

molecules becomes too slow for melting (at the extrapolated, thick short-dashed line in Fig. 2a) to take place on experimental timescales, so that the ice Ih can be ‘supercompressed’. This is supported by my supposition that the glass transition (‘freezing’ of motion in supercooled water) occurs around T_{am} . So in region C and down to around 140 K we can regard the transformation of ice Ih as a supercompression followed by ‘melting’ to a highly viscous liquid, or alternatively, by amorphization to a nearly relaxed HDA ice. Either way, amorphization is clearly not a matter of melting to an equilibrium liquid followed by arrest of molecular motion; rather, it is a transition to an unrelaxed amorphous state which then slowly relaxes further towards a more stable glass-like state. (Below ~ 140 K this relaxation becomes slow on experimental timescales and the HDA ice remains to all intents and purposes unrelaxed). The corollary is that the liquid state (above T_{am}), the viscous liquid or relaxed amorphous state between ~ 160 and ~ 140 K, and the unrelaxed HDA states below ~ 140 K all lie within the same HDA megabasin in configuration space. How this relates to other kinds of amorphization^{12–15} remains an open question at this stage. □

Received 24 April; accepted 8 October 1996.

- Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **310**, 393–395 (1984).
- Wolf, G. H. et al. in *High-Pressure Research: Application to Earth and Planetary Sciences* (eds Syono, Y. & Manghnani, M. H.) 503–517 (Terra Science, Tokyo/Amer. Geophys. Union, Washington DC, 1992).
- Tse, J. S. & Klein, M. L. *J. Chem. Phys.* **92**, 3992–3994 (1990).
- Sciortino, F. et al. *Phys. Rev. E* **52**, 6484–6491 (1995).
- Kanno, H., Speedy, R. J. & Angell, C. A. *Science* **189**, 880–881 (1975).
- Handa, Y. P., Mishima, O. & Whalley, E. *J. Chem. Phys.* **84**, 2766–2770 (1986).
- Johari, G. P., Hallbrucker, A. & Mayer, E. *J. Phys. Chem.* **94**, 1212–1214 (1990).
- Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **314**, 76–78 (1985).
- Mishima, O. *J. Chem. Phys.* **100**, 5910–5912 (1994).
- Angell, C. A. *Science* **267**, 1924–1935 (1995).
- Lüdemann, H.-D. *Ber. Bunsenges. Phys. Chem.* **94**, 325–332 (1990).
- Johnson, W. L. *Prog. Mater. Sci.* **30**, 81–134 (1986).
- Cahn, R. W. & Johnson, W. L. *J. Mater. Res.* **1**, 724–732 (1986).
- Wolf, D., Okamoto, P. R., Yip, S., Lutsko, J. F. & Kluge, M. *J. Mater. Res.* **5**, 286–301 (1990).
- Fecht, H. J. *Mater. Trans. Jpn Inst. Metals* **36**, 777–793 (1995).
- Bertie, J. E., Calvert, L. D. & Whalley, E. *J. Chem. Phys.* **38**, 840–846 (1963).
- Handa, Y. P., Klug, D. D. & Whalley, E. *Can. J. Chem.* **66**, 919–924 (1988).

ACKNOWLEDGEMENTS. I thank T. Grande for his comments, suggestions and discussions, especially on the glass transition temperature of HDA.

CORRESPONDENCE should be addressed to O.M. (e-mail: mishima@nirim.go.jp).

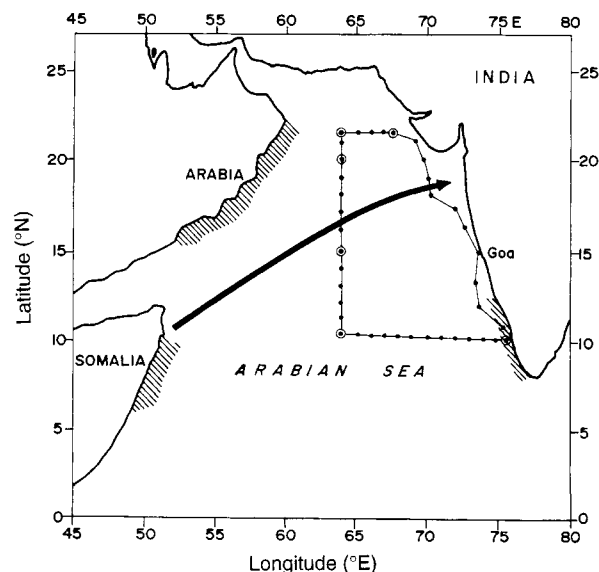


FIG. 1 Cruise track in the Arabian Sea; open circles denote positions where biological data were collected. Temperature, salinity and nutrient profiles were collected from stations shown by both open and filled circles. Hatched areas show regions where coastal upwelling takes place along Somalia, Arabia and the southwestern coast of India during the summer monsoon. The axis of the Findlater Jet, active during this season, is shown by the arrow; northwest of this axis, open ocean upwelling takes place due to positive wind-stress curl and upward Ekman pumping. Temperature, salinity and nutrient profiles were taken at one degree intervals using a Sea-Bird CTD system. Nutrients were analysed with a SKALAR auto-analyser. Primary productivity was measured following ^{14}C clean *in situ* method, chlorophyll *a* was determined with a fluorometer and phytoplankton cells ($>5\mu\text{m}$) were enumerated with an inverted microscope. Zooplankton in the upper mixed layer were collected during the night with a $200\text{-}\mu\text{m}$ mesh; microzooplankton were collected by filtering 20 litres water through a $20\text{-}\mu\text{m}$ mesh. Bacteria and picoplankton were counted using epifluorescence microscopy.

Mechanism of the biological response to winter cooling in the northeastern Arabian Sea

M. Madhupratap*, S. Prasanna Kumar*, P. M. A. Bhattathiri*, M. Dileep Kumar*, S. Raghub Kumar*, K. K. C. Nair† & N. Ramaiah*

* National Institute of Oceanography, Dona Paula, Goa 403 004, India

† Regional Centre of NIO, Cochin 682018, India

THE Arabian Sea is one of the most biologically productive ocean regions¹, mainly due to the upwelling of nutrients during the summer (southwest) monsoon. But the northern Arabian Sea continues to sustain fairly high biological production after the upwelling season and during much of the winter (northeast) monsoon^{2–4}. The processes that enable this high winter productivity have hitherto been poorly understood, being variously attributed to surface cooling effects^{2,3} or wind-driven changes in ocean circulation⁴. Here we present physical, chemical and biological data that indicate that sea surface cooling drives convection processes that lead to the injection of nutrients up into the surface waters of the northeastern Arabian Sea during winter, and that this mechanism of nutrient supply is a dominant control on winter productivity. Observed seasonal changes in bacterial and microzooplankton populations may provide an explanation for the Arabian Sea 'paradox'^{5–8} that mesozooplankton biomass remains more or less invariable throughout the year.

Data were collected in two Joint Global Ocean Flux Study (JGOFS) cruises of ORV *Sagar Kanya* (Fig. 1): 12 April–12 May 1994 (inter-monsoon) and 3 February–5 March 1995 (winter). The physical and chemical conditions during these two cruises were distinctly different. During February 1995, winds were predominantly north/north-easterlies (average speed $\sim 4\text{ m s}^{-1}$) along 64°E . Sea surface temperature (SST) was 27.5°C at 11°N and decreased at a rate of about 0.5°C per degree latitude up to 16°N (Fig. 2). Similarly, air temperature dropped from 26.5°C in the south to $<24^\circ\text{C}$ in the north. The most dramatic change was observed in the mixed-layer depth (MLD) which was thick between 15° and 18°N with a maximum of 120 m at 15°N , but thinned steadily towards the south to 65 m. The salinity structure also showed a deep isohaline layer (100 m thick with >36.4 practical salinity units, psu) in the north which became thinner

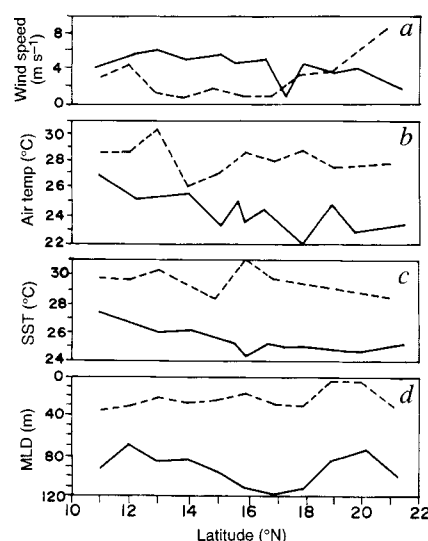


FIG. 2 North-south variations (along 64°E) in wind speed (a), air temperature (b), sea surface temperature (SST; c), and mixed layer depth (MLD; d) during winter (solid line) and the inter-monsoon period (broken line). Measurement uncertainties are $\pm 0.1\text{ m s}^{-1}$, $\pm 0.5^\circ\text{C}$, $\pm 0.5^\circ\text{C}$ and $\pm 1\text{ m}$, respectively. The MLD was defined as the depth where the sea temperature was 1°C less than the SST value.

towards the south. The density (σ_t) structure similarly had a thick isopycnal layer (~ 100 m) of high-density (>24.5) waters in the north which decreased to 22.0 in the south. Along the west coast of India also, the SST decreased from 28 °C in the south to 24.5 °C in the north.

Surface winds along 64° E in April–May 1994 were predominantly northerly and weak (<4 m s⁻¹) up to 17° N, but became progressively stronger (~ 8 m s⁻¹) further north (Fig. 2). There was a marginal decrease in SST from south (28 °C) to north while the air temperature remained more or less constant. The mixed layer was thin (<30 m) despite the comparatively high wind speeds at higher latitudes.

Nitrate in the surface layers is generally below detection levels in the Arabian Sea except during monsoonal upwelling⁹. This is also evident from the inter-monsoon values (Fig. 3b), as well as those data available before winter (see Supplementary Information). In winter, however, the 2 μ M contour shoaled to ~ 30 m at 13° N and surfaced at about 16° N (Fig. 3a).

Increased SST due to increased solar insolation along with weak winds during April–May led to the formation of a highly stratified upper layer with a thin and uniform mixed layer. The reasons for the contrasting deep MLD and high nitrate in the surface layers during winter can be examined in the context of prevailing atmospheric forcings. Under the influence of north/northeasterly winds the possibility of upwelling along the eastern Arabian Sea in winter, by (1) offshore Ekman transport and/or (2) divergence of the surface flow, can be expected. However, the measured wind during winter is too weak to induce any appreciable offshore Ekman transport. Moreover, the deep MLD in the eastern Arabian Sea, both along the western shelf of India as well as along 64° E, does not support this hypothesis. The climate atlas¹⁰ shows a negative windstress curl (-3×10^{-8} dyne cm⁻³) at the southern tip of India in this season which changes to positive north of 10° N. But the magnitude of the positive curl is too small ($<1 \times 10^{-8}$ dyne cm⁻³) to support the co-existence of a deep MLD and high levels of nutrients in the surface layers. To find the driving force, we computed the heat and freshwater (evaporation – precipitation) fluxes along 64° E. This showed that the northern Arabian Sea, especially north of 10° N, experiences a net heat loss (~ 30 W m⁻², Fig. 4a) as well as freshwater loss (~ 125 mm per month, Fig. 4b) during winter. A comparison of climatological specific humidity¹⁰ during October (19 g kg⁻¹) and January (~ 10 g kg⁻¹) in the northeastern Arabian Sea indicates the prevalence of dry air during winter. The cool dry continental

air, brought by the prevailing northeast trade winds, intensify evaporation leading to surface cooling. This starts in December and persists till the end of February. As a result, the northeastern Arabian Sea surface waters experience densification due to excess evaporation over precipitation and turbulent heat loss exceeding the radiative heat gain. The consequent sinking and convection result in the deepening of the mixed layer and an upward transport of nutrients into the surface layers from the base of the mixed layer (80–100 m). A lateral mixing cannot support enhanced nitrate in surface waters as the surrounding waters of the upper layers were nutrient-depleted before winter (see Supplementary Information). Further, the role of shear-driven mixing at the mixed layer base would be small due to the weak wind-driven flow.

A general increase in biological productivity was evident in winter as a result of the changes described above (Table 1). This increase was most apparent in the north and waned towards south, in accord with the observed physical and chemical changes. In the northern Arabian Sea, chlorophyll *a* and primary production values were less than half during inter-monsoon compared to those in winter. They were low in both seasons in the south (Table 1). Similar observations, of higher production during winter (~ 500 – $2,500$ mg C m⁻² d⁻¹) contrasting with lower values during inter-monsoon (~ 200 – 500 mg C m⁻² d⁻¹) have been previously made from the northeastern Arabian Sea^{11,12}. During April–May, a subsurface chlorophyll maximum was distinctly discernible throughout the study area associated with the nutricline; this was less pronounced, particularly in the north, in winter as higher values from the surface (resulting from nutrient injection) progressively declined with depth and lower irradiance (see Supplementary Information). The highest productivity measured in winter (807 mg C m⁻² d⁻¹)—although less than that reported from active upwelling areas in the northwestern Arabian Sea during summer monsoon¹³ (2.5 g C m⁻² d⁻¹)—approaches those reported (900–1,200 mg C m⁻² d⁻¹) during the north Atlantic spring bloom^{14,15}. More importantly, primary production was fairly high over a large stretch of the ocean (north of 15° N), waters which would have otherwise remained oligotrophic as in southern areas. The winter surface chlorophyll *a* concentrations observed (0.2–0.4 mg m⁻³) between 15 and 21° N along 64° E compare well with those obtained from satellite observations³. However, the values inferred from satellite pictures for April–May (0.05–0.3 mg m⁻³) are higher than those found in our study (0.03–0.05 mg m⁻³).

Phytoplankton cell numbers were also very high in the north-

TABLE 1 Biological parameters in the Arabian Sea during winter and inter-monsoon seasons

Parameter	Station (° N, ° E)					
	(21,67)	(21,64)	(19,64)	(15,64)	(11,64)	(10,75)
Chl <i>a</i>	34.4 (10.7)	18.9 (7.6)	20.9 (9.0)	26.7 (16.6)	12.7 (9.2)	9.6 (10.5)
PP	807 (310)	643 (NA)	447 (NA)	606 (168)	335 (163)	(NA) (199)
PC	2–31 (0.08–0.6)	0.3–14 (0.1–2)	1–18 (0.08–2)	0.5–133 (0.1–9)	0.5–3 (0.08–1)	1–5 (0.2–2)
PCP	16–22 (NA)	4–27 (NA)	6–38 (NA)	9–12 (NA)	0.07–0.8 (NA)	0.5–8 (NA)
MZ	0.1–1 (1–4.6)	0.1–0.8 (3.6–15.4)	0.2–2 (4.4–16)	0.3–1.5 (3.8–10.8)	0.4–2.6 (5.4–12)	0.5–1.7 (3–7.6)
TBC	0.03–0.1 (1.1–2.3)	0.006–0.07 (1.2–2)	0.05–0.08 (1.1–1.3)	0.05–0.07 (0.6–2.2)	(NA) (1–2)	(NA) (0.7–1.2)
ZP	86 (40)	64 (66)	50 (NA)	26 (80)	20 (20)	25 (80)

Values given are for February 1995 and April–May 1994 (the latter in parentheses). NA, not available. Parameters are as follows: Chlorophyll *a* (Chl *a*, column value in mg m⁻², upper 120 m); primary production (PP, column value in mg C m⁻² d⁻¹, upper 120 m); phytoplankton cell counts (PC, $\times 10^3$ l⁻¹, range for upper 120 m); picoplankton (PCP, $\times 10^3$ ml⁻¹, range for upper 120 m); microzooplankton counts (MZ, $\times 10^2$ l⁻¹, range for upper 200 m); total bacterial counts (TBC, $\times 10^5$ ml⁻¹, range for upper 100 m); and mesozooplankton biomass (ZP, displacement volume, ml per 100 m³, night values from vertical hauls from mixed layer).

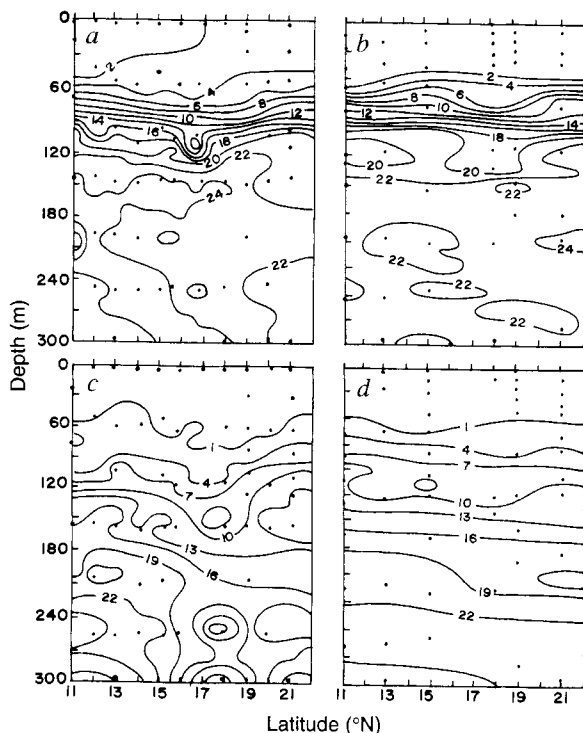


FIG. 3 a, b, Vertical sections along 64° E for nitrate (μM); a, winter and b, inter-monsoon period. c, d, As a, b, but for silicate.

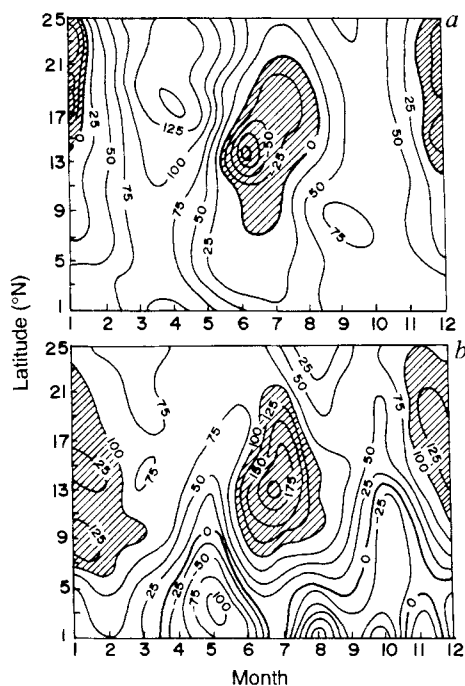


FIG. 4 Time-latitude distribution (month 1 is January) of a, net heat flux (W m^{-2}) and b, evaporation – precipitation (mm per month). The surface heat flux was calculated using bulk aerodynamic formula²¹. Precipitation data were obtained from the very high resolution radiometer (VHRR) on board Indian National Satellite INSAT²²; climatological data on surface meteorological parameters were obtained from the Comprehensive Ocean Atmospheric Data Set (COADS)²³. The hatched region in a represents net heat loss while that in b shows regions where evaporation exceeds precipitation by more than 100 mm per month. Note that there is also a net heat loss, due to evaporation caused by the Findlater Jet, during the southwest monsoon (June–August).

eastern Arabian Sea during winter compared to those in April–May (Table 1). Diatoms accounted for 70–90% in both seasons, dominated by few genera (*Nitzschia*, *Rhizosolenia* and *Chaetoceros*). We attribute the low silicate concentrations in the upper water column in winter (Fig. 3) to rapid uptake by diatoms, because both silicate and nitrate profiles in the northeastern Arabian Sea were nearly similar before winter (see Supplementary Information). Picoplankton (mainly cyanobacteria) abundance in winter was also high, associated with increased nutrients in the north (Table 1).

Two groups of organisms that exhibited trends which were different to those described above were bacteria and microzooplankton. Total bacterial counts were high during inter-monsoon and low during winter. Low bacterial counts ($(0.2\text{--}0.4) \times 10^6 \text{ ml}^{-1}$) before the north Atlantic bloom increased later ($> 2 \times 10^6 \text{ ml}^{-1}$) and remained at this level into early summer (about >1 month after the bloom peaks)^{16,17}, a situation somewhat similar to what we observed in April–May, say 1–2 months after the winter bloom. This increase is probably sustained by the dissolved organic carbon pool that builds up as the bloom grows old. Additionally, we noticed a similar situation during the inter-monsoon period after the summer monsoonal upwelling season (September–October 1993) when the bacterial counts were high ($(0.8\text{--}2.2) \times 10^6 \text{ ml}^{-1}$ in the upper 100 m) between 15° and 21° N along 66° E. Average bacterial carbon during inter-monsoon ($20 \text{ fg C cell}^{-1}$ amounting to $\sim 25 \text{ mg C m}^{-3}$) was about double the highest phytoplankton carbon measured ($1 \text{ mg chlorophyll } a \equiv 30 \text{ mg C}$; concentration of $0.43 \text{ mg Chl } a \text{ m}^{-3} \equiv 12.0 \text{ mg C m}^{-3}$ at 15° N, 64° E at 60 m). In winter, more or less similar values, 10 and 13.8 mg C m^{-3} , respectively, were recorded. Microzooplankton abundances were also higher during inter-monsoon along with bacteria indicating a microbial loop¹⁸.

Recent observations^{5–8} on mesozooplankton from the Arabian Sea seem to indicate that their standing stocks do not vary much despite varying productivity regimes. Our data support this apparent paradox (Table 1, plus our observation that zooplankton biomass in the upper mixed layer was 50–110 ml per 100 m³ in September–October 1993). The community composition was not dissimilar for the two seasons. Fine filter feeders, such as salps and appendicularians, along with small copepods (<2 mm) of the genera *Paracalanus*, *Acrocalanus*, *Clausocalanus*, *Calanus* (minor), *Cosmocalanus* and *Undinula* comprised most of the biomass in both seasons. It is becoming increasingly evident that filter feeders, as well as large copepods, are capable of feeding on microzooplankton or even bacteria^{19,20}. Our results, although preliminary, suggest that the persistence of the high-biomass mesozooplankton community in the northeastern Arabian Sea is made possible through a seitchover from feeding on phytoplankton during productive seasons to the microbial loop during oligotrophic periods. □

Received 15 January; accepted 29 October 1996.

1. Ryther, J. H. & Menzel, D. W. *Deep-Sea Res.* **12**, 199–209 (1965).
2. Banse, K. & McClain, C. R. *Mar. Ecol. Prog. Ser.* **34**, 201–211 (1986).
3. Banse, K. & English, D. C. *Mar. Res.* **2**, 83–103 (1993).
4. Yentsch, C. S. & Phinney, D. A. in *Oceanography of the Indian Ocean* (ed. Desai, B. N.) 57–66 (Oxford & IBH, New Delhi, 1992).
5. Smith, S. L. et al. *Arabian Sea Process Study* (Planning Rep. No. 13, US JGOFS, Woods Hole, Massachusetts, 1991).
6. Madhupratap, M., Handas, P., Ramaiah, N. & Achuthankutty, C. T. in *Oceanography of the Indian Ocean* (ed. Desai, B. N.) 99–112 (Oxford & IBH, New Delhi, 1992).
7. Baars, M. A. et al. in *Monsoons and Pelagic Systems* (ed. Baars, M. A.) 13–34 (Nat'l Museum Nat. Hist., Leiden, 1994).
8. van Couwelaar, M. & Schalk, P. H. in *Arabian Sea Workshop* (abstr.) (Netherlands Inst. for Sea Res., Texel, 1995).
9. Sen Gupta, R. & Naqvi, S. W. A. *Deep-Sea Res.* **31**, 671–706 (1984).
10. Hastenrath, S. & Lamb, P. *Climatic Atlas of the Indian Ocean. Part I, Surface Climate and Atmospheric Circulation* (Univ. Wisconsin Press, 1979).
11. Radhakrishna, K., Devassy, V. P., Bhattathiri, P. M. A. & Bhargava, R. M. S. *Indian J. Mar. Sci.* **7**, 137–139 (1978).
12. Radhakrishna, K., Devassy, V. P., Bhargava, R. M. S. & Bhattathiri, P. M. A. *Indian J. Mar. Sci.* **7**, 271–275 (1978).
13. Owens, N. J. P. et al. *Deep-Sea Res.* **40**, 697–709 (1993).
14. Chapman, D. W., Marra, J. & Takahashi, T. *Deep-Sea Res.* **40**, 151–169 (1993).
15. Harrison, W. J. et al. *Deep-Sea Res.* **40**, 279–305 (1993).
16. Ducklow, H. W., Kirchman, D. L., Quinby, H. L., Carlson, C. A. & Dam, H. G. *Deep-Sea Res.* **40**, 245–263 (1993).

17. Lochte, K., Ducklow, H. W., Fasham, M. J. R. & Stienen, C. *Deep-Sea Res.* **40**, 91–114 (1993).
18. Azam, F. *et al. Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
19. Christoffersen, K., Riemann, B., Hansen, L. R., Klysner, A. & Sørensen, H. B. *Microb. Ecol.* **20**, 253–272 (1990).
20. Painting, S. J., Moloney, C. L. & Lucas, M. I. *Mar. Ecol. Prog. Ser.* **100**, 55–69 (1993).
21. Bunker, A. F. *Mon. Weath. Rev.* **104**, 1122–1140 (1970).
22. Ramesh Kumar, M. R. & Prasad, T. G. *J. Geophys. Res.* (in the press).
23. Woodruff, S. D., Sultz, R. J., Jenne, R. L. & Steuer, P. M. *Bull. Am. Meteorol. Soc.* **68**, 1239–1250 (1987).

SUPPLEMENTARY INFORMATION is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

ACKNOWLEDGEMENTS. We thank a number of our colleagues for help in collection and processing data; P. N. Aravindakshan, T. C. Gopalakrishnan, P. Haridas, C. Raghukumar, S. Mohanraju, S. Sawant, M. Gauns, A. Rajendran, S. N. DeSousa, S. Sardesai, V. V. S. S. Sarma, A. Almeida and M. T. Babu. This work was supported by the Department of Ocean Development, New Delhi.

CORRESPONDENCE should be addressed to M.M. (e-mail: madhu@csnio.ren.nic.in).

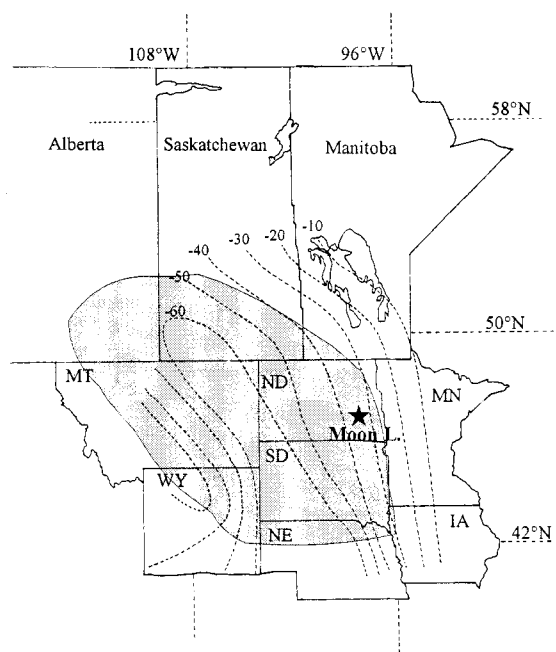


FIG. 1 Map of the Northern Great Plains (shaded area) showing the location of Moon Lake (46° 51' 27" N, 98° 09' 30" W). Moon Lake⁸ is 35 ha in size with a maximum depth of ~13 m. Isolines are equal precipitation minus evapotranspiration (cm yr^{-1}). Figure is modified from Fritz *et al.*¹³.

Greater drought intensity and frequency before AD 1200 in the Northern Great Plains, USA

Kathleen R. Laird^{*†}, Sherilyn C. Fritz[‡], Kirk A. Maasch[§] & Brian F. Cumming^{||}

^{*} Department of Ecology, Evolution and Behavior and the Limnological Research Center, University of Minnesota, Minneapolis, Minnesota 55455, USA

[‡] Department of Earth and Environmental Sciences, Lehigh University, Bethlehem, Pennsylvania 18015, USA

[§] Department of Geological Sciences and Institute for Quaternary Studies, University of Maine, Orono, Maine 04469, USA

^{||} Paleoeological Environmental Assessment and Research Lab (PEARL), Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

EXTREME large-scale droughts in North America, such as the 'Dust Bowl' of the 1930s, have been infrequent events within the documented history of the past few hundred years, yet this record may not be representative of long-term patterns of natural variation of drought intensity and frequency. In the Great Plains region of central North America, historical droughts have persisted longer than in any other part of the United States¹, but no detailed records of drought patterns in this region have hitherto been obtained that extend beyond the past 500 years. Here we present a reconstruction of drought intensity and frequency over the past 2,300 years in the Northern Great Plains, based on lake salinity fluctuations inferred from fossil diatom assemblages. This record, of sub-decadal resolution, suggests that extreme droughts of greater intensity than that of the 1930s were more frequent before AD 1200. This high frequency of extreme droughts persisted for centuries, and was most pronounced during AD 200–370, AD 700–850 and AD 1000–1200. We suggest that before AD 1200, the atmospheric circulation anomalies that produce drought today were more frequent and persistent.

The present-day climate of the Northern Great Plains (NGP, shaded area in Fig. 1) is highly variable as a result of its continental location and the changing influence of major air masses of differing characteristics: (1) warm dry flow from the Pacific, (2) cold dry air from the Arctic, and (3) moist tropical air from the Gulf of Mexico². Most (75%) of the annual precipitation in the NGP falls during the growing season from April to September³ and decreases across this region from ~520 mm in the northeast to

350 mm in the southwest (Fig. 1). Before agricultural settlement the NGP was a vast region of prairie and today forms one of the main agricultural belts of North America.

Spring and summer precipitation in the NGP is largely a result of the southerly flow of warm moist air from the Gulf of Mexico and local convective precipitation. During periods of drought, moisture from this tropical airmass is blocked by strong anticyclonic (high-pressure) activity in the upper troposphere over the north-central United States. This high-pressure system over the NGP is usually associated with other upper-level anticyclonic conditions over the east-central Pacific and the North Atlantic, and strong low-pressure troughs off both coasts^{4,5}. Convective precipitation also decreases when there is a lack of moisture from local land sources⁶, thus amplifying the moisture deficit.

Here we reconstruct long-term variability in the intensity, frequency and duration of droughts in the NGP over the past 2,300 years from the diatom-inferred salinity record of Moon Lake, a topographically closed-basin lake in eastern North Dakota (Fig. 1). Fluctuations in salinity reflect changes in the hydrological budget of the lake as a result of changes in effective moisture (precipitation minus evapotranspiration). Our working assumption is that periods of positive water balance (precipitation > evapotranspiration) are reflected by higher lake levels and lower salinities, whereas when the water balance is negative, lake levels are lower and salinity higher⁷. Comparison of the diatom-inferred salinity of Moon Lake for the past 100 years with measured precipitation minus evapotranspiration suggests that Moon Lake has this type of hydrochemical response to fluctuations of effective moisture⁸.

Diatoms are common members of the algal flora of inland saline and freshwater lakes and their distribution is highly related to lakewater salinity^{9–13}. Weighted-averaging regression techniques¹⁴ can be used to estimate salinity optima of diatom taxa from the analysis of modern diatom assemblages in surficial sediments of lakes spanning a salinity gradient. The estimated optima can then be used to infer salinity values from assemblages in cores using transfer functions. The predictive ability of diatom-based salinity transfer functions is strong and highly significant as judged

[†] Present address: Paleoeological Environmental Assessment and Research Lab (PEARL), Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada.

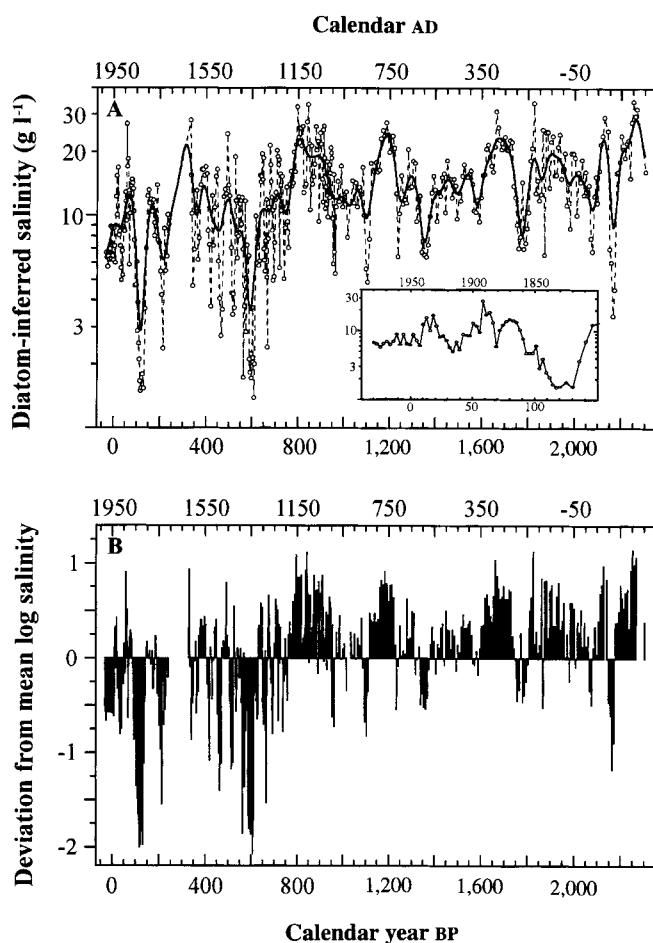


FIG. 2 a, Dotted line is the diatom-inferred salinity in Moon Lake presented on a logarithmic scale owing to the logarithm based derivation of the transfer function. Average temporal resolution between samples is 5.3 years. The solid line is a fast Fourier transformation of 5-year equal interval interpolated data using a 10-point smoothing window. Inset, expanded-scale plot of salinity changes over the past two centuries. b, Deviation from the mean log salinity of the past 2,300 years. Mean log salinity is 2.4 (11.1 g l^{-1}). The 92-year gap from 240 to 332 years BP is due to desiccation of the top 16-cm of core section two. Salinity reconstructions are based on the relative abundance of diatom taxa that were present in the upper 4 m of an 11.8-m long core collected from $\sim 13 \text{ m}$ depth from the deep basin of Moon Lake¹⁵. Approximately 45 diatom taxa (see Supplementary Information for stratigraphies) are predominant in the core¹⁵. These include: the saline taxa, *Chaetoceros elmorei/muelleri* and *Cyclotella quillensis*; hyposaline taxa, *Cyclotella meneghiniana*, *Rhoicosphenia curvata* and *Cocconeis placentula* v. *euglypta*; and freshwater taxa, *Cymbella cymbiformis* v. *nonpunctata*, *Stephanodiscus minutulus* and *S. parvus*. Only 3 samples of 427 were eliminated due to preservation problems. Salinity estimates were calculated from a transfer function developed by Fritz *et al.*^{12,13} based on estimated salinity optima by weighted-average regression of diatom taxa from their present-day distributions in 53 lakes in the NGP that span a salinity gradient from 0.65 to 270 g l^{-1} . Chronology is based on four AMS dates calibrated to calendar years (Table 1) and stratigraphic correlation of distinct peaks in diatom species composition in the top metre of the long core with a ^{210}Pb -dated 1.2-m short core¹⁵.

from their coefficient of determinations, which range from ~ 0.8 to 0.9 (refs 10, 11, 13). Analogue analyses for the past 2,300 years indicate that $>90\%$ of the core samples from Moon Lake are well represented by the modern diatom assemblages used in the transfer function¹⁵; the transfer functions therefore permit reliable estimates of past lakewater salinity.

Major historical droughts (for example, in the 1930s and 1890s) are clearly evident in the Moon Lake record (Fig. 2a). In addition, a freshwater interval (salinity $< 3 \text{ g l}^{-1}$) during AD 1820–35 corresponds to both an episode of abruptly cold temperatures frequent in many records in the early to mid-1800s¹⁶, and to a drought-free period from 1825 to 1838 inferred from tree-ring records on the periphery of the US Great Plains¹⁷.

Over a longer timescale, the inferred salinity of Moon Lake suggests that extreme droughts greater or equal in intensity to those of the 1930s were more frequent before AD 1200 (Fig. 2). This high frequency of extreme drought persisted for centuries, and was most pronounced during AD 200–370, AD 700–850 and AD 1000–1200. Recent tree-ring data from the western USA^{18–20} do not obviously indicate this major switch in the frequency of extreme droughts despite adequate temporal resolution of two of the records^{18,20}), but do suggest that intervals of severe multi-decadal drought have been common during the past few millennia. Correspondence of persistently dry intervals at Moon Lake with a high frequency of extreme drought during AD 250–350 and AD 700–850 reconstructed from giant sequoia (*Sequoiadendron giganteum* (Lindley) Buchholz) tree-ring records in the San Joaquin Valley, California¹⁸, suggests a common forcing by

large-scale circulation. The Moon Lake drought interval from AD 1000 to AD 1200 is coeval with the 'medieval warm period' (MWP), an interval of warmer temperatures from about AD 1000 to AD 1300²¹ whose geographical extent, timing and climatic manifestation is much debated²². In the Sierra Nevada region evidence based on drowned tree stumps¹⁹ and tree-ring records from subalpine conifers²⁰ also indicate extreme droughts during medieval time during AD 900–1100 and AD 1200–1350, and during AD 1020–1070, AD 1200–1220 and AD 1250–1350, respectively.

The intervals of the lower inferred salinity since AD 1200 suggests that the past ~ 750 years has been wetter/cooler than any of the preceding $\sim 1,500$ years (Fig. 2b). This pronounced shift

TABLE 1 Moon Lake radiocarbon dates

Core depth (cm)	Material	^{14}C date* ($\pm 1 \text{ s.d.}$)	Calendar years BP (and 1σ range)
110–114	Charcoal flakes	460 ± 60	509 (477–530)
177–192	Charcoal flakes	830 ± 60	725 (672–782)
245–246	Charcoal flakes	$1,080 \pm 50$	966 (936–1,055)
342–347	Charcoal flakes	$2,300 \pm 60$	2,334 (2,206–2,349)

The first column shows the depth in core from which charcoal was collected for dating. Radiocarbon dates (years BP) $\pm$ error (1 standard deviation in years BP) are indicated. Radiocarbon dates were calibrated to calendar year with the CALIB 3.0 program using the bidecadal calibration curve³¹. There was no ambiguity in the conversion from radiocarbon years to calendar years. Calendar years BP are indicated with (in parentheses) minimum and maximum 1σ calendar-age ranges³¹ obtained from intercepts.

* Determined by accelerator mass spectrometry (AMS) at the Lawrence Livermore Laboratory.

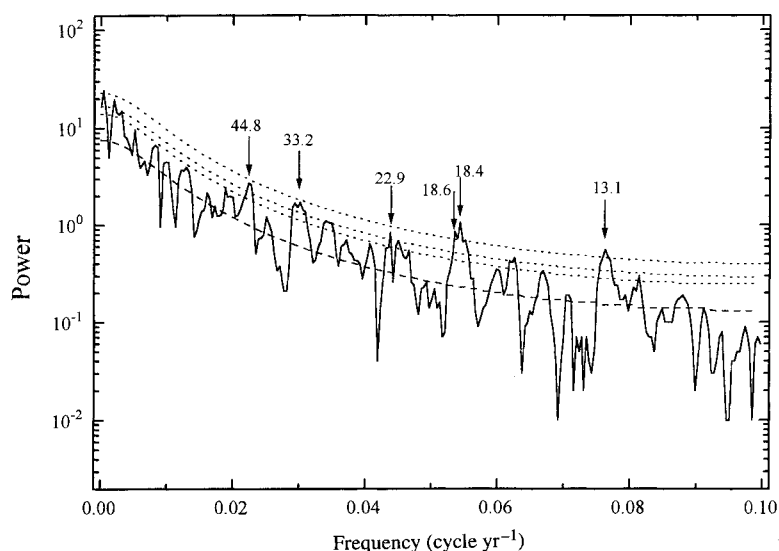


FIG. 3 Spectral analysis of the salinity curve interpolated at 5-year intervals using the multi-taper method³². A set of tapers were applied to the data in the time domain before Fourier transformation. By choosing discrete prolate spheroidal sequences (slepian tapers) the resistance to spectral leakage is optimized producing a maximally band-limited spectral estimate. The multi-taper method allows for an estimate of a narrowband F -test for the presence of significant lines in the spectrum. The significance of broader peaks, possibly quasi-periodic components, is assessed by comparing the spectrum with the red noise background (long-dashed curve) and 90, 95, and 99% confidence levels (dashed curves) estimated using the robust procedure of Mann and Lees³³. Periodicities that fall within the 95 and 99% confidence levels are indicated by arrows.

of mean salinity around AD 1200 coincides with the end of the MWP and the onset of the 'little ice age' (LIA) in North America (AD 1300–1850)²³. It is also coeval with a distinct change in the $^{14}\text{CO}_2$ record from low $^{14}\text{CO}_2$ production during the medieval solar activity maximum (AD 1100–1250) to 650 years of higher $^{14}\text{CO}_2$ production as a result of reduced solar activity²⁴.

Although the LIA has often been characterized as a continuously cold interval of global extent, recent syntheses suggest it was much more complex, with decade-long cold episodes separated by warmer intervals, and with variable behaviour between regions²⁵. The Moon Lake record provides evidence for a hydrologically complex LIA interval, with wet periods interspersed with short episodes of drier conditions, at times comparable to the droughts of the 1930s. The two major freshwater periods (salinity $< 3 \text{ g l}^{-1}$) at Moon Lake (Fig. 2a) are contemporaneous with periods of reduced fire frequency in Minnesota (AD 1240–1440 and since AD 1600), inferred from charcoal accumulation from varved sediments in Elk Lake²⁶, and the intervening period of greater fire frequency (AD 1440–1600) coincides with a higher incidence of saline conditions. These records together suggest a regionally complex LIA interval.

Spectral analysis of the inferred salinity curve for the past 2,300 years (Fig. 3) suggests that the recurrence of drought may be associated with solar and lunar cycles. Spectral peaks with significance levels $>99\%$ occur at 18.6, 18.4 and 13.1 years. The 18.6-year lunar periodicity has been found in previous drought records inferred from tree-ring data in the interior and western North America²⁷. Peaks with 95% significance levels include 22.9, 33.2 and 44.8 years. These oscillations may be related to the 22-year Hale magnetic solar cycle and to harmonics of the 11-year sunspot amplitude cycle.

Climate reconstructions at Moon Lake for the past 750 years and a 500-year tree-ring record of summer precipitation in Alberta on the western edge of the NGP²⁸ indicate that the frequency of drought was not appreciably different from the past 100 years. In sharp contrast, the longer-term Moon Lake record suggests that extreme droughts were more frequent before 750 years before present (AD 1200). This suggests that the pattern of a persistent upper-level high-pressure ridge over central North America, accompanied by other upper-level highs and lows across the continent and adjacent oceans which have been associated with summer drought in the Plains in the historic period^{4,5}, may have been more frequent in the past. The correspondence of the extreme drought intervals at Moon Lake with the few other high-resolution records from western North America for the past two millennia suggests that these anomalies were probably the result of large-scale circulation patterns, and thus that the linked atmo-

spheric-oceanic system, possibly in concert with solar variations, was characteristically in a different mode before AD 1200.

Current climate models suggest that increasing atmospheric CO_2 concentrations could cause an intensification of the extremes of the hydrological cycle^{29,30}, thereby increasing the frequency and severity of floods and droughts. Current models, however, are inadequate to predict reliably the changes in specific regions. Nonetheless, given the adverse effects of the relatively low frequency of severe droughts in recent history, a return to the earlier mode of climate variability in the NGP before AD 1200 would be devastating. □

Received 9 April; accepted 24 October 1996.

1. Karl, T. J. *J. Clim. Appl. Meteorol.* **22**, 1356–1366 (1983).
2. Bryson, R. *Geogr. Bull.* **8**, 228–269 (1966).
3. Oladipo, E. J. *Climatol.* **5**, 655–664 (1985).
4. Trenberth, K., Branstator, G. & Arkin, P. *Science* **242**, 1640–1645 (1988).
5. Namias, J. J. *J. Clim. Appl. Meteorol.* **22**, 30–39 (1983).
6. Chang, F. & Wallace, J. *Mon. Weath. Rev.* **115**, 1253–1269 (1987).
7. Fritz, S. *Limnol. Oceanogr.* **35**, 1771–1781 (1990).
8. Laird, K., Fritz, S., Grimm, E. & Mueller, P. *Limnol. Oceanogr.* **41**(5), 890–902 (1996).
9. Cumming, B., Wilson, S., Hall, R. & Smol, J. *Bibliotheca Diatomologica* **31**, 1–207 (1995).
10. Gasse, F., Juggins, S. & Khelifa, L. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **117**, 31–54 (1995).
11. Wilson, S., Cumming, B. & Smol, J. *J. Paleolimnol.* **12**, 197–221 (1994).
12. Fritz, S., Juggins, S., Battarbee, R. & Engstrom, D. *Nature* **352**, 706–708 (1991).
13. Fritz, S., Juggins, S. & Battarbee, R. *Can. J. Fish. Aquat. Sci.* **50**, 1844–1856 (1993).
14. Line, J., ter Braak, C. & Birks, H. J. *Paleolimnol.* **10**, 147–152 (1994).
15. Laird, K. thesis, Univ. Minnesota (1996).
16. Jacoby, G., Ivanciu, I. & Ulan, L. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **64**, 69–78 (1988).
17. Stockton, C. & Meko, D. J. *J. Clim. Appl. Meteorol.* **22**, 17–29 (1983).
18. Hughes, M. & Brown, P. *Clim. Dynam.* **6**, 161–167 (1992).
19. Stine, S. *Nature* **369**, 546–549 (1994).
20. Graumlich, L. *Quat. Res.* **39**, 249–255 (1993).
21. Lamb, H. *Climate: Present, Past, and Future Vol. 2, Climatic History and the Future* (Methuen, London, 1977).
22. Hughes, M. & Diaz, H. *Clim. Change* **26**, 109–142 (1994).
23. Porter, S. *Quat. Res.* **26**, 27–48 (1986).
24. Jirikovic, J. & Damon, P. *Clim. Change* **26**, 309–316 (1994).
25. Bradley, R. & Jones, P. (eds) *Climate Since A.D. 1500* (Routledge, London, 1992).
26. Clark, J. *Nature* **334**, 233–235 (1988).
27. Currie, R. J. *Geophys. Res.* **89**, 1295–1308 (1984).
28. Case, R. & MacDonald, G. *Quat. Res.* **44**, 267–275 (1995).
29. Hansen, J. et al. in *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations* (ed. Schlesinger, M.) 211–229 (Elsevier Science, Amsterdam, 1991).
30. Tsionis, A. *Nature* **382**, 700–702 (1996).
31. Stuiver, M. & Reimer, P. *Radiocarbon* **35**, 215–230 (1993).
32. Thompson, D. *IEEE Proc. Natl. Aerosp. Electron. Conf.* **70**, 1055–1096 (1982).
33. Mann, M. & Lees, J. *Clim. Change* **33**, 409–445 (1996).

SUPPLEMENTARY INFORMATION is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

ACKNOWLEDGEMENTS. We thank D. Engstrom for providing the ^{210}Pb dating of the short core and E. Grimm for collection and AMS dating of the long core, and J. P. Smol, P. Glaser, H. E. Wright, Jr, R. Forester and R. Battarbee for comments on the manuscript. This work was supported by the US National Science Foundation, NOAA and a Research Training Grant from University of Minnesota (K.R.L.).

CORRESPONDENCE should be addressed to K.R.L. (e-mail: lairdk@biology.queensu.ca).

Fossils of large terrestrial arthropods from the Lower Devonian of Canada

William A. Shear*, Patricia G. Gensel†
& Andrew J. Jeram‡

* Department of Biology, Hampden-Sydney College, Hampden-Sydney, Virginia 23943, USA

† Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

‡ Ulster Museum, Botanic Gardens, Belfast BT9 5AB, UK

TERRESTRIAL arthropods have a fossil record that reaches back to at least the Upper Silurian (420 million years ago). Most available data on these early land animals comes from a few sites at which abundant microarthropod remains are preserved as organic cuticle fragments. Such cuticle assemblages imply that early terrestrial ecosystems were dominated by small arthropods of the soil and litter. We describe here new impression and impression/compression fossils from the Emsian stage of New Brunswick and Québec, Canada, which include millipeds, arthropleurids, and a scorpion with book-lungs preserved. They suggest that surface macroarthropods were also a significant component of the land fauna.

Until the 1980s, most of our knowledge of early terrestrial arthropods came from two sites of Lower Devonian age: Rhynie, near Aberdeen in Scotland¹ (Pragian), and Alken-an-der-Mosel in Germany^{2,3} (Emsian). At Rhynie, arachnids are found preserved among *in situ* vegetation by the rapid precipitation of silica during eruptive phases of a hot spring complex⁴. The arthropods are found by searching thin sections or translucent chips of chert. The Alken deposit is slightly younger than Rhynie, and the arthropods are preserved as impressions on the shale matrix, without any preserved cuticle. Rare, mainly fragmentary, remains of large terrestrial arthropods have been found in the Old Red Sandstone (ORS; Lower Devonian) of Britain⁵. Most of these isolated occurrences are also impressions lacking preserved cuticle.

Since 1984, most evidence of early terrestrial arthropods has come from dispersed cuticle sites, from which specimens are recovered by bulk maceration of cuticle-bearing sediments. The arthropod cuticle remains are organically preserved and typically very small, from 0.1 to 1.0 mm in length. Faunas from Gilboa, New York, USA⁶⁻⁹ (Givetian, Middle Devonian), and Ludford Corner, Shropshire, England¹⁰ (Pridoli, Upper Silurian), include a range of arachnids, centipeds, arthropleurids and possibly insects. The ecomorphotypic composition of these assemblages is comparable with modern litter or soil communities dominated by detritivores and their predators. This view of early terrestrial arthropod communities is also consistent with the evidence from the Rhynie Chert. However, the Old Red Sandstone impression fossils differ in faunal composition and ecotypic morphology. They represent non-centiped myriapods⁵, and scorpions that reached nearly 1 m in length¹¹.

The new fossils from Canada described here are essentially impression fossils, but some specimens also have organic cuticle preserved. Thus the Emsian sediments of New Brunswick and Québec may eventually yield both impression and organically preserved arthropod material, bridging the gap between impression assemblages (ORS and Alken) and cuticle assemblages (Ludford Lane, Rhynie, Gilboa). This may provide a less biased and more complete sample of the terrestrial arthropod community than obtained elsewhere, accompanied by a well-studied terrestrial flora. In addition, these fossils represent the earliest fauna of terrestrial animals yet reported from North America.

A rich flora of zosterophyllophytes, trimerophytes, lycopsids and several other plant types has been described from localities

near Dalhousie Junction in New Brunswick and the Gaspé peninsula, Québec, Canada¹². The Campbellton Formation (New Brunswick) is Emsian, perhaps late early Emsian in age, based on spore assemblage correlations (D. C. McGregor, personal communication). The Battery Point Formation (Québec) ranges from early Emsian to early Eifelian. On the basis of spore assemblages analysed from each plant-bearing locality in these sequences, the Gaspé fossils discussed here are slightly older than those from New Brunswick, although they collectively fall within the 'Annulatus-sextantii' assemblage zone of Richardson and McGregor¹³.

The terrestrial animal remains from Gaspé discussed here were obtained in 1980 by W. and P. Gensel from locality 'W' of Gensel and Andrews¹⁴, in the lower part of Battery Point Formation. They are preserved in fine-grained, grey to grey-green sandstone/siltstone, which may have been part of a sandy alluvial plain. A single specimen of a supposed archaeognath insect has been recovered from macerations of the Battery Point Formation¹⁵, but this specimen appears not to meet generally agreed upon criteria for proof of its fossil nature¹⁰ and may be a recent contaminant. Permineralized plant stems in this formation show evidence of wounding that may have been caused by arthropods with chewing and sucking mouthparts¹⁶.

Four detailed impression fossils of what would appear to be the same species of milliped (Fig. 1) have been found at locality W within the Battery Point Formation. None of the four shows any trace of original cuticle. The complete specimens are about 45 mm in length and 4–5 mm wide. The segments were probably originally cylindrical, divided into an anterior prozonite and a posterior metazonite, as in Carboniferous and modern forms. The metazonite bears fine, oblique striations, similar to those found on the much later (Upper Pennsylvanian) xyloiidulids¹⁷. Specimens preserved in lateral view show numerous slender legs, 2–3 mm long. Examination at higher magnifications shows that there are two pairs per segment, a definitive character of millipeds. In life, this animal would have closely resembled the extant spirobolid milliped *Narceus annularis*. We suggest that these Emsian millipeds pushed their way among the decayed stems lying on the ground, probably feeding on them as well, or burrowed in soft soil, based on comparison with extant milliped ecomorphotypes.

Arthropod remains were found at two localities of volcanic-derived fluvial sediments in New Brunswick, locality F, and a site we call locality T, which may be close to or identical to locality B of Gensel and Andrews¹⁸. Preservation of these animal remains

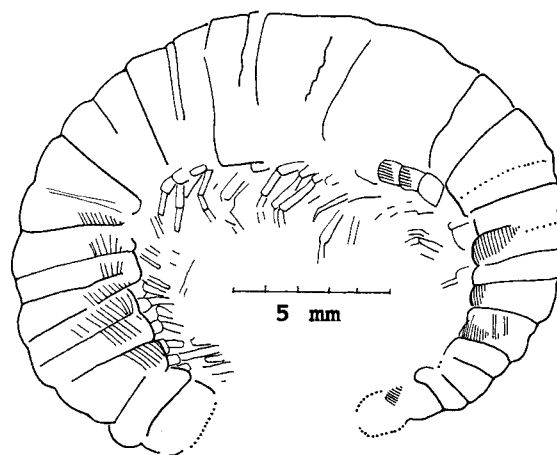


FIG. 1 Camera lucida drawing of a fossil milliped from the Battery Point Formation, Québec, Canada. The anterior end of the headless fossil is to the right, based on the overlap of prozonites by the preceding metazonites. Three additional partial specimens appear to represent the same species.

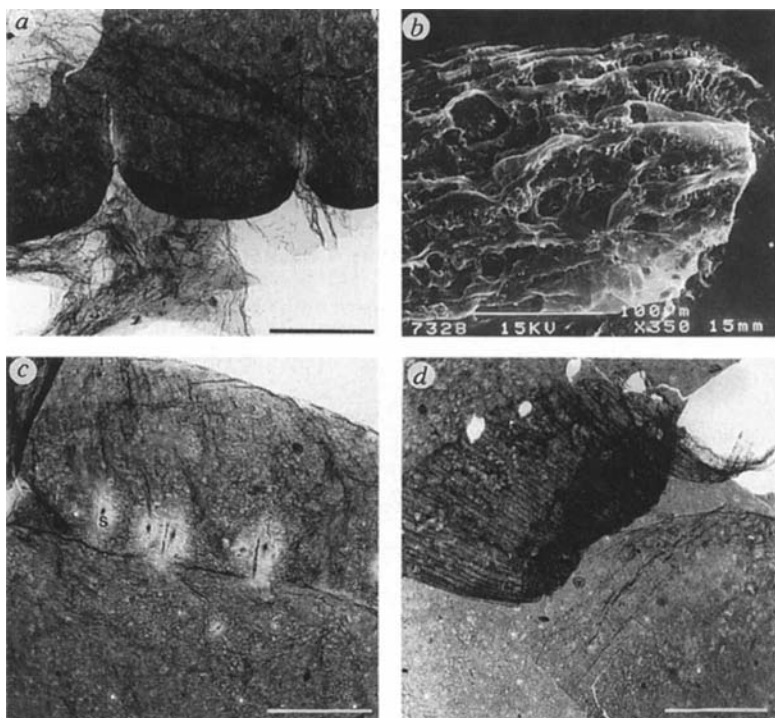


FIG. 2 Photomicrographs of mesoscorpion cuticle from the Campbellton Formation, New Brunswick, Canada. *a*, The unique scalloped posterior margin of a ventral abdominal plate, with thin intersegmental cuticle attached. *b*, Scanning electron micrograph of book-lung tissue; note marginal strengthening bars at the atrial end of each lamella and the reticulate thickenings of the lamellae themselves. *c*, Anterior margin of a ventral abdominal plate, showing the anterior transverse ridge, and five sensory organs (*s*) similar to slit sensilla. *d*, Book-lung (dark mass above) attached to thin sternite cuticle and abdominal plate. Scale bars in *a*, *c* and *d*, 1 mm; photomicrographs were taken at a magnification of $\times 4$.

ranges from impressions only, to impressions with coalified cuticle, to impressions/compressions with organically preserved cuticles. Two fossils from locality F are parts of midbody segments of a species of the eoarthroleurid genus *Eoarthroleura*, which was first found at the approximately contemporaneous Alken-an-der-Mosel site in Germany¹⁹. The genus is also known from the Upper Silurian Ludford Lane deposit, and a later (Upper Devonian, Famennian?) deposit at South Mountain, Schoharie County, New York²⁰. Trackways recently reported²¹ from the Upper Silurian Clam Bank Formation of Newfoundland, Canada, may be attributable to *Eoarthroleura*. On the basis of the size of the Canadian segments, it is reasonable to assume a width of up to 2.5 cm and a length of 15–20 cm. Little is known or can be inferred about the habits of eoarthroleurids, an entirely extinct group. The better-known Carboniferous *Arthropleura* evidently fed on soft pith material from the interior of decaying aborescent lycopod logs²², a niche not available to these earlier arthropleurids.

Parts of a single large scorpion, represented by organically preserved cuticle, have been found at locality T (=B?) in the Campbellton formation. Impressions of three distorted abdominal plates, with some adherent cuticle, were visible on the rock surface. Allowing for distortion and incompleteness, these plates have an estimated width of 17 mm and length of 5–7.5 mm (anterior plates are usually shorter than posterior ones); by comparison with other early fossil scorpions, this suggests an overall length of 9–9.5 cm, excluding pedipalps.

The sample containing scorpion parts was macerated in hydrofluoric acid, destroying the impression, but yielding ~ 120 cuticle fragments, the largest with an area of 25 mm². The cuticle, derived exclusively from the ventral abdominal surface, includes recognizable portions of abdominal plate anterior, lateral and posterior margins (Fig. 2*a*), and the overlying sternites. Lamellate structures clearly represent book-lungs, providing the earliest direct evidence of air-breathing in scorpions. Although the material is fragmentary, portions of the book-lungs connected to sternite and abdominal plate (Fig. 2*d*) by very thin arthrotridial membrane indicate that the gross morphology is consistent with the respiratory structures of the Lower Carboniferous scorpion *Pulmonoscorpius*²³. Unlike *Pulmonoscorpius*, in which the respiratory surfaces of the book-lung lamellae are not known in detail,

the New Brunswick material has lamellae in an excellent state of preservation, with a strengthening bar of cuticle at the posterior (atrial) end of each, and reticulate thickening on the lamellar surfaces similar to extant buthid scorpions (Fig. 2*b*). There is no doubt that this material is derived from a scorpion because the abdominal plates have a typical anterior transverse ridge (Fig. 2*c*) and the tiny conical setae that are characteristic of late Silurian and early Devonian primitive mesoscorpions. Early mesoscorpions were first considered aquatic²⁴, although leg morphology and the presence of a rudimentary oral tube are more consistent with terrestrial life²⁵. The presence of book-lungs in the Campbellton Formation scorpion supports the hypothesis that all of the mesoscorpions, including the largest, were indeed part of the terrestrial fauna²⁶.

The most obvious difference between these Canadian sites and most others yielding Siluro–Devonian terrestrial arthropods is the presence of the remains of large animals (10–20 cm long) and fossils of millipeds. Millipeds are among the earliest known land animals but they have not been found except as isolated occurrences in Upper Silurian and Lower Devonian sediments of Britain⁵. Identifiable milliped cuticle has never been found at any of the cuticle-producing sites; millipeds are much more likely to occur as calcified remains or as impression fossils, because their cuticle, impregnated with calcium carbonate, has a low preservation potential in the strongly reducing environments that preserve cuticle. Although scorpion cuticle has a relatively high preservation potential, the remains of large scorpions may be missed because they are likely to be so fragmented as to be unrecognizable or recognizable only as scorpion cuticle. Work is needed on the relative preservation potentials of terrestrial arthropod cuticles, recognizing that each of the major components of cuticle, protein, lipid, chitin and minerals (absent from the cuticles of most non-crustacean arthropods), and the manner in which they are organized within cuticles, responds distinctively to biogenic attack and sediment pH and chemistry²⁷.

In reflected light, preserved cuticle of microarthropods lacks any distinctive appearance, and would be most unlikely to be detected by an examination of rock surfaces; impression fossils of macroarthropods buried within macerated samples or unseen on the surface would be destroyed by the maceration process that

yields cuticle. The connection of cuticle fragments with impression fossils of euarthropods has now enabled us to recognize patches of previously unassigned cuticle as belonging to larger individuals.

The arthropod communities at Gilboa, Rhynie and Ludford Lane more closely resemble, in their species composition and the size of the animals involved, modern soil and litter communities, dominated by small detritivores and their predators^{1,6-10}. It now appears that this may be at least in part due to preservational style and the unusual conditions required organically to preserve arthropod cuticle. Alken and scattered Old Red Sandstone sites feature impression fossils of larger, surface-dwelling macroarthropods, and cuticle is absent. The taphonomic significance of the

Canadian sites discussed above is that both impression fossils and organically preserved cuticle of macroarthropods are present. What is sampled at these sites is a fauna of surface-dwelling or burrowing macroarthropods, albeit also composed of detritivores (millipeds and euarthropods) and carnivores (scorpions). Evidence of possible wounding of plants in the Battery Point Formation by arthropods suggests that true herbivores as well may have been a part of the ecological scheme at this time¹⁶. Further work at these sites may increase taxonomic diversity, size range and habitat type, helping to clarify the debate on the roles of herbivory and detritivory in the earliest terrestrial ecosystems and the establishment of trophic structures on land. □

Received 28 August; accepted 29 October 1996.

- Hirst, S. *Annu. Mag. Nat. Sci.* **9**, 455–474 (1923).
- Størmer, L. *Senckenberg. Leth.* **51**, 335–369 (1970).
- Brauckmann, C. *Jber. naturwiss. Ver. Wuppertal* **47**, 168–173 (1994).
- Trewin, N. *Trans. R. Soc. Edinb.* **84**, 433–442 (1994).
- Almond, J. E. *Phil. Trans. R. Soc. Lond. B* **309**, 227–238 (1985).
- Shear, W. A. et al. *Science* **224**, 492–494 (1984).
- Shear, W. A., Selden, P. A., Bonamo, P. & Grierson, J. D. *Am. Mus. Nov.* **2901**, 1–74 (1987).
- Shear, W. A. & Bonamo, P. M. *Am. Mus. Nov.* **2927**, 1–30 (1988).
- Shear, W. A., Schawaller, W. & Bonamo, P. M. *Nature* **341**, 527–529 (1989).
- Jeram, A. J., Selden, P. A. & Edwards, D. *Science* **250**, 658–661 (1990).
- Shear, W. A. & Kukalová-Peck, J. *Can. J. Zool.* **68**, 1807–1834 (1990).
- Gensel, P. G. *Proc. 3rd N. Amer. Paleontol. Conv.* **1**, 199–204 (1982).
- Richardson, J. B. & McGregor, D. C. *Geol. Surv. Can. Bull.* **363**, 1–79 (1986).
- Gensel, P. G. & Andrews, H. N. *Plant Life in the Devonian* (Praeger, New York, 1984).
- Labandeira, C., Beall, B. S. & Hueber, F. M. *Science* **242**, 913–916 (1988).
- Banks, H. P. & Colthart, B. J. *Am. J. Bot.* **80**, 992–1001 (1993).
- Hoffman, R. L. in *Treatise on Invertebrate Paleontology, Part R, Arthropoda 2* (ed. Moore, R. C.)

- 572–606 (University of Kansas Press, Lawrence, 1969).
- Andrews, H. N., Gensel, P. G. & Forbes, W. H. *Palaeontology* **17**, 387–408 (1974).
- Størmer, L. *Senckenberg. Leth.* **57**, 87–183 (1976).
- Shear, W. A. & Selden, P. E. N. *Jb. Geol. Paläont. Abh.* **196**, 347–375 (1995).
- Wright, J. L., Quinn, L., Briggs, D. E. G. & Williams, S. H. *Can. J. Earth Sci.* **32**, 304–313 (1995).
- Rolfe, W. D. I. & Ingham, J. K. *Scott. J. Geol.* **3**, 118–124 (1967).
- Jeram, A. J. *Nature* **343**, 360–361 (1990).
- Kjelleswig-Waering, E. N. *Palaeontogr. Am.* **55**, 1–287 (1986).
- Selden, P. A. & Jeram, A. J. *Trans. R. Soc. Edinb.* **80**, 303–310 (1989).
- Jeram, A. J. *Palaeontology* **37**, 513–550 (1994).
- Plotnick, R. E. in *Arthropod Paleobiology* (ed. Mikulic, D. G.) 177–196 (1990).

ACKNOWLEDGEMENTS. We thank J. Dunlop and P. Manning for Fig. 2b. This work was supported by grants from the NSF (W.A.S., and A.J.J., P.G.G. and B.S.R.) and by the George Frederick Matthew Research Grant of the New Brunswick Museum, and the George Cooley Fund of the University of North Carolina.

CORRESPONDENCE should be addressed to W.A.S. The fossils described here will be deposited in the New Brunswick Museum after detailed descriptions have been published elsewhere.

Photorespiration protects C3 plants from photooxidation

Akiko Kozaki* & Go Takeba†

* Laboratory of Applied Botany, Department of Agriculture Biology, Kyoto University, Kyoto 606, Japan

† Laboratory of Applied Biology, Department of Life Science, Kyoto Prefectural University, Kyoto 606, Japan

PLANTS absorb light for photosynthesis but as light can itself be dangerous to plants, they need to protect themselves against its damaging effects. Here we show that photorespiration can act as such a defence mechanism. We constructed transgenic tobacco plants enriched or reduced in plastidic glutamine synthetase (GS2), a key enzyme in photorespiration. Those transgenic plants having twice the normal amount of GS2 had an improved capacity for photorespiration and an increased tolerance to high-intensity light, whereas those with a reduced amount of GS2 had a diminished capacity for photorespiration and were photoinhibited more severely by high-intensity light compared with control plants. We conclude that photorespiration protects C3 plants from photoinhibition.

Photorespiration is the metabolism of phosphoglycolate produced by the oxygenase activity of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)^{1,2}. Photorespiration lowers the energy efficiency of photosynthesis, particularly in C3 plants¹, which include major crops such as wheat and rice. However, attempts to reduce rates of photorespiration in order to increase net photosynthesis have all failed^{1,2}.

One suggestion³ is that photorespiration is effective in dissipating excess photochemical energy. A variety of photorespiratory mutants of the C3 plant *Arabidopsis thaliana* are able to grow in a high-CO₂ environment in which plants do not need to photorespire, but such plants cannot grow in normal air in which photorespiration can occur^{1,4}. These observations indicate that photorespiration is necessary for the survival of C3 plants in air^{1,2}, but the physiological function of photorespiration remains unclear.

We reasoned that if photorespiration operates as a mechanism for the dissipation of excess light energy, a plant with a large photorespiratory capacity should withstand high-intensity light better than a plant with a limited capacity. We therefore estimated changes in the levels of photorespiratory metabolites in order to determine the rate-limiting step in photorespiratory metabolism. Under photorespiration conditions, ammonia started to accumulate and the concentrations of glutamine and serine then gradually increased in the leaves of the plant (Fig. 1), indicating that reassimilation of ammonia must be the rate-limiting step of the metabolic pathway, as proposed previously^{5,6}.

We constructed transgenic tobacco plants with high levels of GS2 and a large capacity for photorespiration by using rice GS2 complementary DNA⁷ in *Agrobacterium*-mediated transformation. About 30 tobacco clones carrying rice GS2 cDNA were

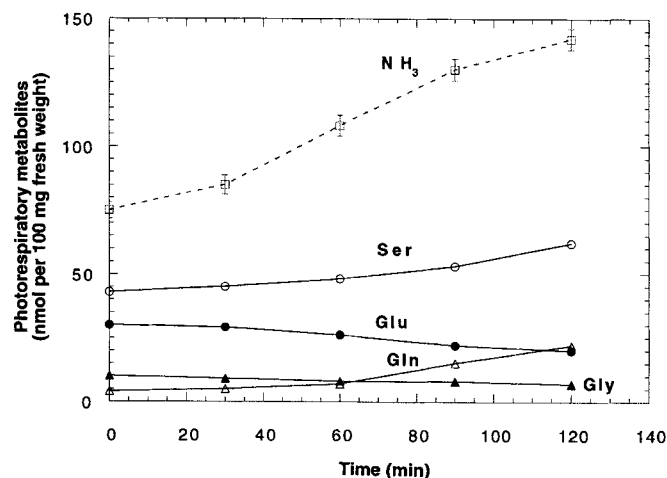


FIG. 1 Changes in photorespiratory metabolites during high-intensity light irradiation ($1,500 \mu\text{mol m}^{-2} \text{s}^{-1}$) in the absence of CO₂ in tobacco leaves. The amount of phosphoglycolate, glycolate or glyoxylate was less than 2 nmol per 100 mg fresh weight, and none of these acids accumulated during the timescale shown.

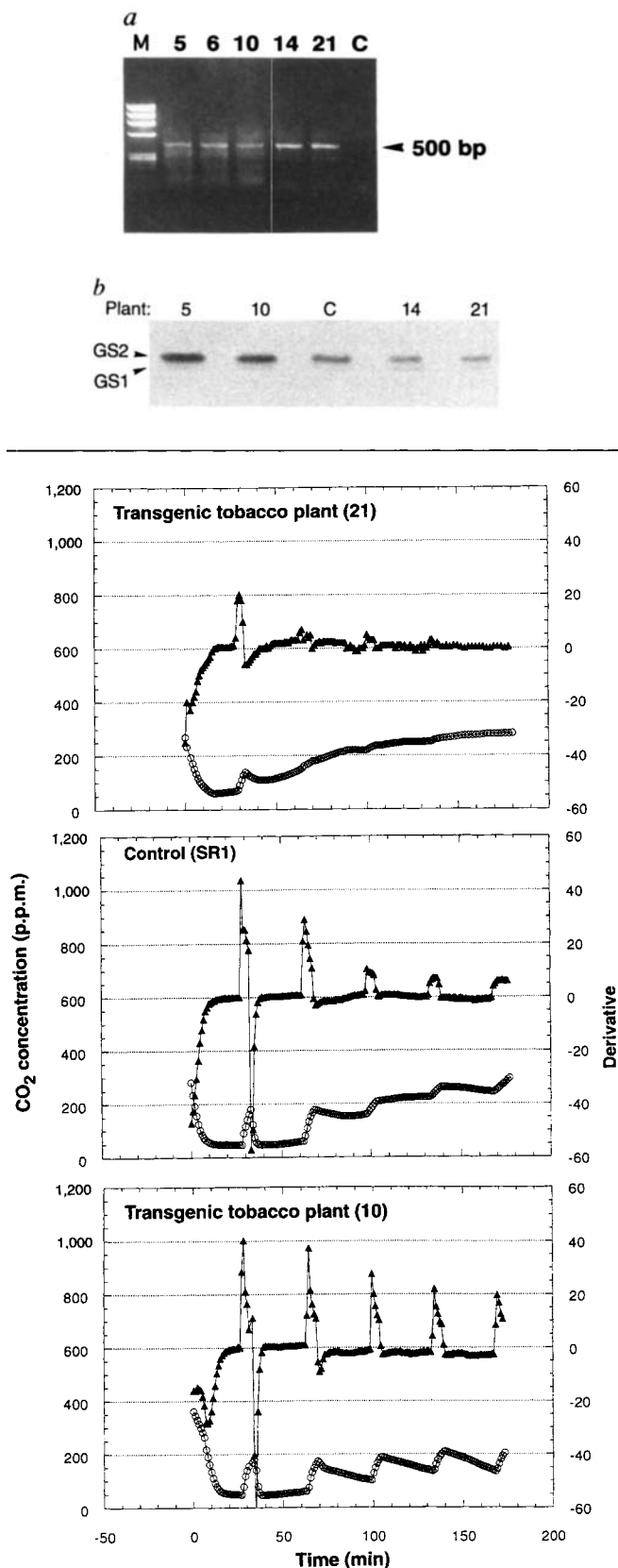


FIG. 3 Estimation of the capacity for photorespiration by leaves of transgenic and control tobacco plants. The leaf was irradiated for 30 min in an elliptical clear plastic box and left in the dark for 5 min; the cycle was repeated. The concentration of CO_2 (circles) in the closed space (flow rate of the gas: 120 ml min^{-1}) was traced by an LI-COR CO_2 analyser (LI-COR, Lincoln, USA). The difference in rate (p.p.m. CO_2 per min) over a minute is represented by triangles. The CO_2 burst in the dark gives an estimate of photorespiration². Temperature, 25°C .

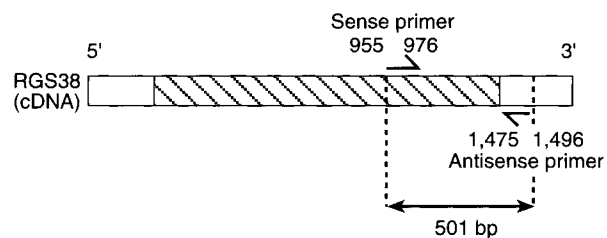


FIG. 2 a, Detection of rice GS2 cDNA in transgenic tobacco plants by PCR analysis²⁵. M, DNA size markers (ΦX174 *Hae*III digest); C, control, non-transformed tobacco plant. Numbers represent transgenic tobacco plant clones. Two oligonucleotides, corresponding to nucleotide positions 955–976 and 1,475–1,496 in the cDNA of rice GS2 (ref. 7), were used as sense and antisense primers, respectively. b, Detection of rice GS2 in transgenic plants by western blotting^{26,27}. Proteins ($10 \mu\text{g}$) in the leaves of the R_2 generation were separated by SDS-PAGE. GS activity in the leaves of transgenic plant 5 was about 2.5 times higher than in the control (data not shown), which was confirmed by separating the homogenate on a DEAE–Sephacel column as described²⁸.

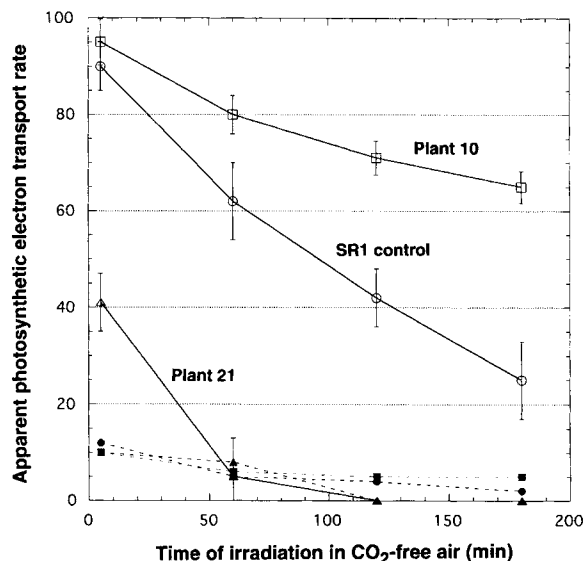


FIG. 4 Changes in the electron transport rate (ETR; $\mu\text{mol electrons m}^{-2} \text{ s}^{-1}$) in transgenic tobacco plant leaves (squares, 10; circles, SR1 control; triangles, 21) under high-intensity light ($1,300 \mu\text{mol m}^{-2} \text{ s}^{-1}$) irradiation in CO_2 -free air (solid line: $21\% \text{ O}_2$, dashed line: $2\% \text{ O}_2$). ETR were estimated using PAM-2000 fluorometer (Heinz Waltz, Effeltrich, Germany)²⁹. Before the experiment, leaves were exposed to low-intensity light ($200 \mu\text{mol m}^{-2} \text{ s}^{-1}$) for 30 min in CO_2 -free, $21\% \text{ O}_2$.

selected by polymerase chain reaction (PCR) (Fig. 2a) and the amount of GS2 in the leaves of the R_2 generation was determined by western blotting (Fig. 2b). There was more GS2 in some clones (numbers 5 and 10) than in control plants (SR1), but less than expected in others (numbers 6, 14 and 21), perhaps because of co-suppression⁸.

We compared the photorespiratory capacity of our transgenic tobacco plants with a wild-type control burst that occurs in the dark immediately after transfer from the light². Under the conditions used in Fig. 3, the rate of photorespiration in a control tobacco leaf decreased rapidly, but continued for much longer in a transgenic one (number 10; Fig. 3). But in another transgenic leaf (number 21) with a reduced level of GS2 (Fig. 2b), photorespiration was very low and diminished rapidly (Fig. 3, top panel), demonstrating the key role of GS2 in photorespiration⁶.

Figure 4 shows the changes in electron transport rate (ETR) during photosynthesis in leaves of transgenic tobacco plants induced by intensive radiation with light ($1,300 \mu\text{mol m}^{-2} \text{ s}^{-1}$) in CO_2 -free air. At the start of this irradiation, the ETR in normal air

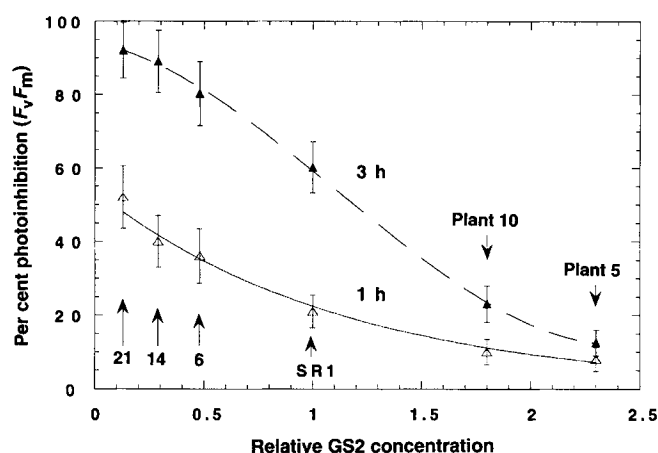


FIG. 5 Rates of photoinhibition in transgenic tobacco plant leaves, estimated by the decrease in chlorophyll fluorescence. Fluorescence was measured in mature leaves using a PAM-2000 fluorometer after the leaves had been dark-adapted for 15 min. The decrease in F_v/F_m (ref. 29) for 1 h or 3 h after irradiation with high-intensity light represents the photoinhibition.



FIG. 6 Comparison of photooxidation damage in a transgenic tobacco plant (5) leaf enriched for GS2 (right) and a control leaf (left) after exposure to high-intensity light ($2,000 \mu\text{mol m}^{-2} \text{s}^{-1}$) for 24 h. A sample from each leaf was punched out before irradiation for the assay of GS.

(450 p.p.m. $\text{CO}_2/21\% \text{O}_2/79\% \text{N}_2$) was 118, 105 and $50 \mu\text{mol electrons m}^{-2} \text{s}^{-1}$ in plants 10, SR1 and 21, respectively. The difference in the ETR in 21% and 2% O_2 may be due to photorespiration^{3,9}. We found that the contribution of photorespiration to the ETR at the start of high-intensity light irradiation was high in CO_2 -free air, indicating that photorespiration can drive electron transport even in the absence of photosynthesis. ETRs decreased in all plants during intensive light irradiation, but the decrease in transgenic tobacco plant 10, which had a larger capacity for photorespiration (Fig. 3), was smaller than in plants SR1 and 21: the ETR in the leaves of plant 21, which had only a limited capacity for photorespiration (Fig. 3), fell almost to zero after irradiation for 1 h. Figure 5 shows the relation between GS2 levels in the leaves of transgenic plants and the photoinhibition produced by intensive irradiation with light for 1–3 hours in the apparatus used for Fig. 3. We found that photoinhibition in the leaves of transgenic plants containing less GS2 (plant numbers 21, 14, 6) was more severe than in the control (SR1), but was milder in plants with more GS2 (plants 10 and 5). We conclude that the amount of GS2 in the leaves of the transgenic tobacco plants correlates well with their capacity for photorespiration (Fig. 3) and with their tolerance to high-intensity light (Fig. 5).

To visualize the effect of high-intensity light on photooxidation of tobacco leaves, we compared the degradation of chlorophyll in a leaf from transgenic tobacco plant 5 with that in a control SR1 leaf, in a closed box. Figure 5 shows that the control leaf was severely damaged by high-intensity light, whereas one from plant 5 was unaffected. The accumulation of ammonia was slowed down in the transgenic leaves (data not shown), indicating that these plants had an increased capacity for recycling the ammonia produced in photorespiration. We also found that chlorophyll was degraded much faster in the leaves of those transgenic plants with less GS2 (plants 21, 14 and 6) than in the non-transgenic leaves of SR1 (data not shown).

The idea of a photoprotective role for photorespiration^{3,9,10} is controversial^{11–14}, but is supported by our results (Figs 4–6). Figure 7 depicts how such a scheme might work. When sufficient CO_2 is supplied to a leaf, ATP and NADPH are used in the Calvin–Benson cycle to produce sugars, and photorespiration does not occur. However, in air, especially when the supply of CO_2 is limited, phosphoglycolate is formed by the oxygenase reaction of Rubisco; this phosphoglycolate may be metabolized via photorespiration to produce CO_2 and phosphoglycerate, which would drive the Calvin–Benson cycle (Fig. 4). *In vivo* photorespiratory (C2) metabolism forms the united (C2/C3) cycle with

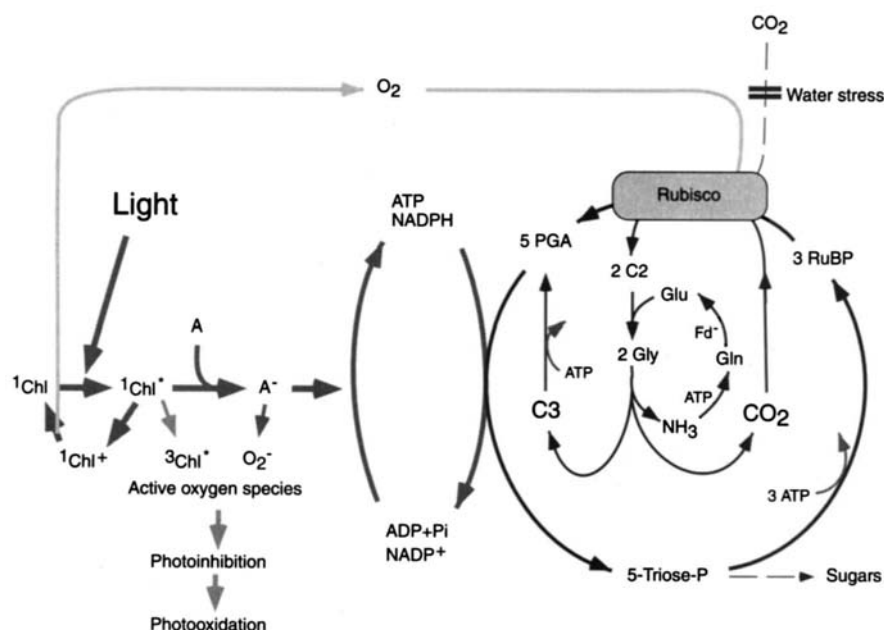


FIG. 7 The proposed photoprotective role of photorespiration. Abbreviations: ^1Chl , singlet state of chlorophyll; ^3Chl , triplet state of chlorophyll; A^- , excited state of chlorophyll; A , electron carrier, for example plastoquinone, cytochrome complex or ferredoxin; PGA, 3-phosphoglycerate; RuBP, ribulose-1,5-bisphosphate; Fd^- , ferredoxin; C2, two-carbon compounds (such as phosphoglycolate, glycolate or glyoxylate); C3, three-carbon compounds (such as serine, hydroxypyruvate or glycinate).

the Calvin–Benson (C3) cycle¹⁵. When the supply of CO₂ from outside the leaves is cut off by closure of the stomata and the C2 cycle does not work either, the C3 cycle will not operate, so light energy may be diverted to produce active oxygen, which could damage the photosynthetic apparatus^{12,16}. This may explain why plants suffer severe photoinhibition when photorespiration is suppressed. An important function of photorespiration may be to derive the Calvin–Benson cycle when the supply of CO₂ is limited, a new role for photorespiration.

Under high-intensity light, recycling of the ammonia produced by the glycine decarboxylase complex may be the rate-limiting step (Fig. 1). If there is sufficient GS2 to recycle the ammonia, photorespiration may drive the Calvin–Benson cycle and protect plants from photooxidation (Figs 4–6). However, when there is insufficient GS2 and high-intensity light is supplied continuously, ammonia would accumulate and the amino donor would be depleted¹⁷, which could slow down the Calvin–Benson cycle (that is, the ETR) and lead to photoinhibition (Fig. 5). Rapid photoinhibition is known to occur when the chloroplast NADP system is completely reduced^{9,18}.

Most algae and all cyanobacteria have a CO₂ pump to concentrate CO₂ around Rubisco; the glycolate formed is metabolized or excreted into the surrounding medium^{2,19}. Submerged aquatic plants in unshaded shallow water and land plants have to live under higher-intensity light and lower levels of CO₂ than those experienced by most algae or cyanobacteria, and may have developed photorespiration in order to drive the Calvin–Benson cycle under such conditions. □

Methods

Mature young leaves were used for all experiments. The amount of ammonia was estimated by using aeration and Nessler's reagent²⁰. Amino-acid concentrations were estimated according to ref. 21. A *Sma*–*Sac*II fragment containing a *GUS* gene in pBI121 (ref. 22) was removed and replaced by rice GS2 cDNA⁷ (RGS38, in the sense orientation) to give 35S-GS2 cDNA. Ti-binary plasmids were mobilized from *Escherichia coli* to *Agrobacterium tumefaciens* LBA4404 by triparental mating²³. The *Agrobacterium*-mediated transformation of tobacco was carried out by the leaf disc method²⁴.

Received 2 September; accepted 29 October 1996.

- Ogren, W. L. *Annu. Rev. Plant Physiol.* **35**, 415–442 (1984).
- Husic, D. W., Husic, H. D. & Tolbert, N. E. *CRC Crit. Rev. Plant Sci.* **5**, 45–100 (1987).
- Powles, S. B. & Osmond, C. B. *Plant Physiol.* **64**, 982–988 (1979).
- Somerville, C. R. *Annu. Rev. Plant Physiol.* **37**, 467–507 (1986).
- Häusler, R. E., Blackwell, R. D., Lea, P. J. & Leegood, R. C. *Planta* **194**, 406–417 (1994).
- Wallisgrove, R. M., Turner, J. C., Hall, N. P., Kendall, A. C. & Bright, S. W. *Plant Physiol.* **83**, 155–158 (1987).
- Sakamoto, A. et al. *Plant Mol. Biol.* **13**, 611–614 (1989).
- Matzke, M. A. & Matzke, J. M. *Plant Physiol.* **107**, 679–685 (1995).
- Wu, J., Neimanis, S. & Heber, U. *Bot. Acta* **104**, 283–291 (1991).
- Heber, U., Bligny, R., Streb, P. & Douce, R. *Bot. Acta* **109**, 307–315 (1996).
- Demmig-Adams, B. & Adams, W. W. *Annu. Rev. Plant Physiol.* **43**, 599–626 (1992).
- Aro, E. M., Vergin, I. & Anderson, B. *Biochim. Biophys. Acta* **1143**, 113–134 (1993).
- Long, S. P. & Humphries, S. *Annu. Rev. Plant Physiol.* **45**, 633–662 (1994).
- Brestic, M., Comic, G., Fryer, M. J. & Baker, N. R. *Plant* **196**, 450–457 (1995).
- Tolbert, N. E. in *Regulation of Atmospheric CO₂ and O₂ Photosynthetic Carbon Metabolism* (eds Tolbert, N. E. & Preiss, J.) 8–33 (Oxford Univ. Press, Oxford, 1994).
- Asada, K. in *Causes of Photooxidative Stress and Amelioration of Defence Systems in Plants* (eds Foyer, C. H. & Mullineaux, P. M.) 77–104 (CRC, London, 1994).
- Wendler, C., Putzer, A. & Wild, A. J. *Plant Physiol.* **139**, 666–671 (1992).
- Krause, G. H. & Berrend, U. *FEBS Lett.* **200**, 298–302 (1986).
- Powles, S. B. *Annu. Rev. Plant Physiol.* **35**, 15–44 (1984).
- Sumner, J. B. in *The Enzymes* (eds Sumner, J. B. & Myrback, K. M.) Vol. 1, 873–892 (Academic, New York, 1951).
- Sato, K. et al. *J. Agric. Food Chem.* **40**, 808–810 (1991).
- Jefferson, R. A., Kavanagh, T. A. & Bevan, M. W. *EMBO J.* **6**, 3901–3907 (1987).
- Bevan, M. *Nucleic Acids Res.* **12**, 8711–8721 (1984).
- Horsch, R. B. et al. *Science* **227**, 1229–1231 (1985).
- Edwards, K., Johnstone, C. & Thompson, C. *Nucleic Acids Res.* **19**, 1349–1350 (1991).
- Blake, M., Johnson, K. H., Russell-Jones, G. J. & Gotchich, E. C. *Analyt. Chem.* **136**, 175–179 (1984).
- Sakamoto, A., Takeba, G. & Tanaka, K. *Plant Cell Physiol.* **31**, 677–682 (1990).
- Takeba, G. *Plant Cell Physiol.* **24**, 1469–1476 (1983).
- Genty, B. E., Briantais, J. M. & Baker, N. R. *Biochim. Biophys. Acta* **990**, 87–92 (1989).

ACKNOWLEDGEMENTS. We thank N. Murata, K. Asada and K. Satou for discussions. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas from Ministry of Education, Science and Culture, Japan.

CORRESPONDENCE and requests for materials should be addressed to G.T. (e-mail: takeba@chiko.kpu.ac.jp).

Sensitization of meningeal sensory neurons and the origin of headaches

A. M. Strassman*†, S. A. Raymond†‡ & R. Burstein*†§

* Department of Anesthesia and Critical Care, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts 02215, USA

† Department of Anaesthesia, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Department of Anesthesia, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA

§ Department of Neurobiology, Harvard Institute of Medicine, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA

THE headaches that accompany certain intracranial pathologies (such as meningitis, subarachnoid haemorrhage and tumour) have been considered to result from mechanical or chemical stimulation of pain-sensitive structures of the intracranial meninges^{1,2}. Although the recurrent headache of migraine is of unknown origin and is not accompanied by an identifiable pathology, it shares with intracranial headaches features that suggest an exaggerated intracranial mechanosensitivity (worsening of the pain by coughing, breath-holding or sudden head movement^{1,3}). One possible basis for such symptoms would be a sensitization of meningeal afferents to mechanical stimuli. Previous studies of neuronal responses to meningeal stimulation have focused primarily on cells in the central portion of the trigeminal pathway, and have not investigated the possible occurrence of sensitization^{4–12}. We have recorded the activity of primary afferent neurons in the rat trigeminal ganglion that innervate the dural venous sinuses. Chemical stimulation of their dural receptive fields with inflammatory mediators both directly excited the neurons and enhanced their mechanical sensitivity, such that they were strongly activated by mechanical stimuli that initially had evoked little or no response. These properties of meningeal afferents (chemosensitivity and sensitization) may contribute to the intracranial mechanical hypersensitivity that is characteristic of some types of clinically occurring headaches, and may also contribute to the throbbing pain of migraine.

Anatomical studies in man and animals have shown that major blood vessels of the intracranial meninges, including the dural venous sinuses, receive a sensory innervation from the trigeminal nerve which originates primarily from cells in the medial (ophthalmic) part of the trigeminal ganglion^{13–15}. Direct stimulation of these vessels in neurosurgical patients can evoke painful sensations, which some investigators have described as headache-like, and which are typically referred to the ophthalmic region of the trigeminal dermatome¹. We used single-unit extracellular recording techniques to study the sensory properties of cells in the rat trigeminal ganglion that supply this intracranial meningeal innervation.

Trigeminal ganglion neurons with A-delta and C-fibre axons were identified as meningeal afferents by their response to electrical stimulation of the dura overlying the ipsilateral transverse sinus (Fig. 1a, b). The transverse sinus, which is a venous space between the inner and outer layers of the intracranial dura, is supplied by the tentorial nerve^{13,14}, and was found to be a particularly well innervated region in our previous studies of central *fos* expression evoked by dural stimulation¹⁶. The majority of neurons (33/45) that were activated by dural shock exhibited mechanosensitive receptive fields on the dura, from which they could be activated by punctate probing with von Frey hairs or by stroking with a blunt rod (Fig. 1c, d). Receptive fields were restricted to the dura overlying and immediately adjacent to the transverse sinus, and did not extend across the midline. Mechanosensitivity was found more frequently in neurons with C-fibres

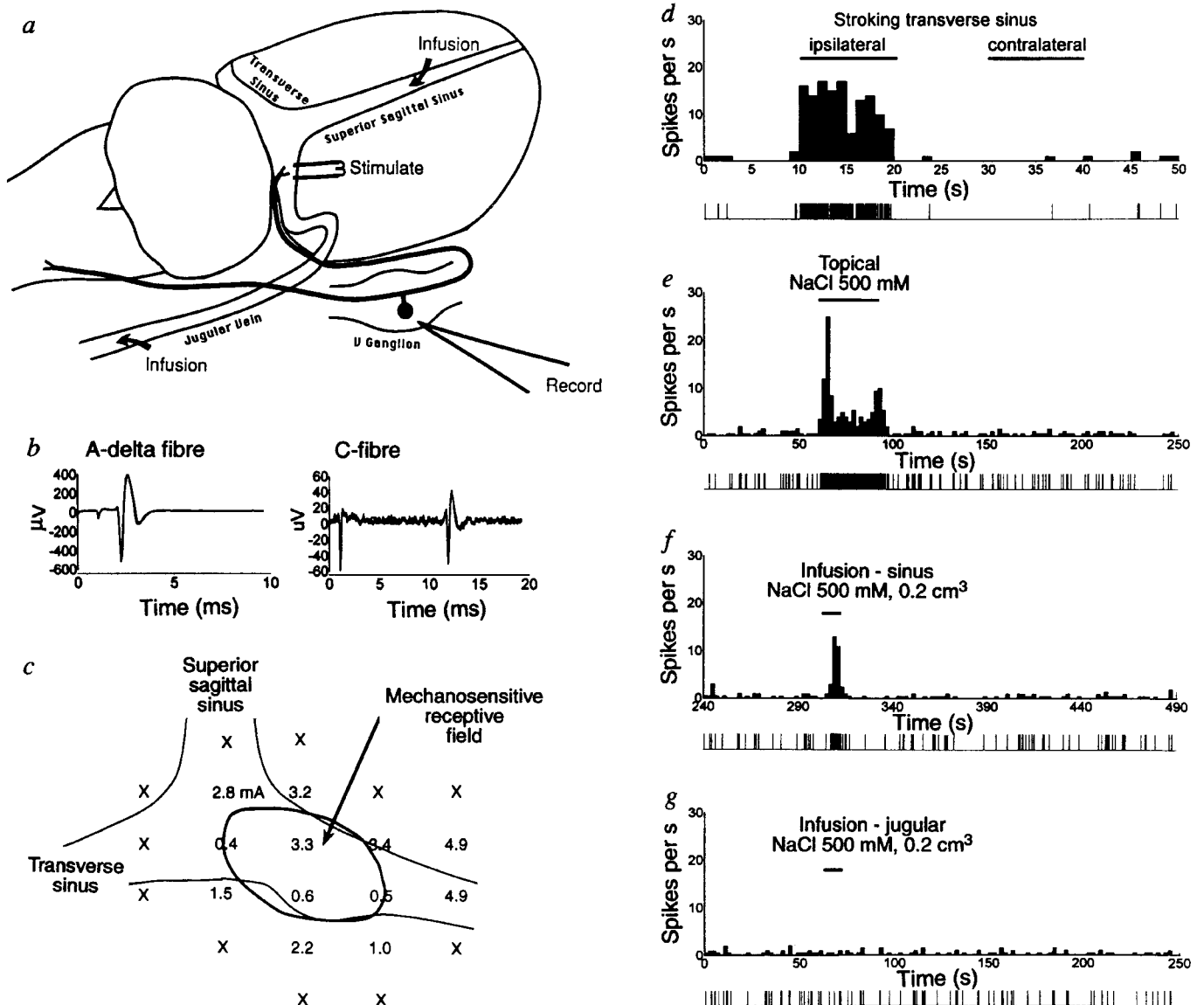


FIG. 1 **a**, Experimental setup. **b**, Responses of two trigeminal ganglion cells to single-shock stimulation of the dura overlying the ipsilateral transverse sinus. Each trace shows 3 superimposed sweeps. Response latencies indicate that the axons of these cells conduct in the A-delta and C-fibre range, respectively (12.5 ms^{-1} and 1.2 ms^{-1} , respectively) based on a conduction distance measured at 12.5 mm . Conduction velocities (c.v.) of cells that responded to dural shock ranged from 0.26 to 20.8 ms^{-1} . The mean c.v.s were 11.6 ± 4.1 for fast A-delta fibres, 3.5 ± 0.89 for slow A-delta fibres, and 0.76 ± 0.38 for C-fibres. **c**, Map of the response thresholds (mA) of a trigeminal ganglion cell (c.v. 4.8 ms^{-1}) at different dural

stimulation sites on and around the ipsilateral transverse sinus. The area from which a response to stroking could be evoked is also indicated by the boundary line. **d**, Histogram of discharge of the cell shown in **c** during stroking of the dura over the ipsilateral transverse sinus in the bounded area, and during stroking of the corresponding contralateral dura. **e–g**, Responses of the cell shown in **c** and **d** to chemical stimulation (500 mM NaCl in SIF) delivered either by topical application to the dural surface of the transverse sinus (**e**) or by infusion into the superior sagittal sinus (**f**), and absence of response to control infusion delivered into the jugular vein (**g**) (see 7Methods for further explanation).

($\leq 1.5 \text{ ms}^{-1}$) (13/15) and slowly conducting A-delta fibres ($1.5 >$ conduction velocities (c.v.) $< 5 \text{ ms}^{-1}$) (10/10) than in neurons with rapidly conducting A-delta fibres ($5\text{--}21 \text{ ms}^{-1}$) (10/20) ($P < 0.01$, chi-square). Among 26 mechanosensitive neurons that were tested with von Frey monofilaments, mechanical thresholds ranged from 0.03 to $>4.0 \text{ g}$ (monofilaments stiffer than 4.0 g were not usually tested, to avoid puncturing the dura). Among neurons whose thresholds could be determined ($\leq 4.0 \text{ g}$; $n = 23$), the mean threshold was $0.86 \pm 0.98 \text{ g}$ (mean $\pm$ s.d.), with no significant differences between neurons with fast A-delta fibres ($1.8 \pm 2.0 \text{ g}$; $n = 5$), slow A-delta fibres ($0.78 \pm 0.97 \text{ g}$; $n = 8$) and C-fibres ($1.0 \pm 1.1 \text{ g}$; $n = 10$).

A subset of the neurons that responded to dural shock (18/22 mechanosensitive and 1/5 mechanically insensitive neurons) also responded to topical application of alginate or inflammatory agents

to the dura (9/11 C-fibres, 7/8 slow A-delta and 3/8 fast A-delta fibres). Responses were found to hypertonic sodium chloride (400 mM, 12/18), potassium chloride (150 mM, 13/21), pH-neutral buffer solutions of low (10/16) or high (4/5) osmolarity, capsaicin (3/9), and a soup of inflammatory agents (histamine, serotonin, bradykinin, and prostaglandin E_2 , pH 5) (10/18) (Fig. 2a–d). Responsiveness to these 'chemical stimuli' (inflammatory, osmotic (low-osmolarity buffer), and ionic (KCl and NaCl)) usually occurred together. Of the 15 neurons that were tested with each of these 4 agents, 7 responded to all, 5 responded to none, and 3 responded to all except the inflammatory soup. Nearly all (18/19) of the neurons that could be activated by chemical stimuli were also mechanosensitive.

Five of the neurons that responded to topically applied chemical agents were also tested for responses to intravascular infusion

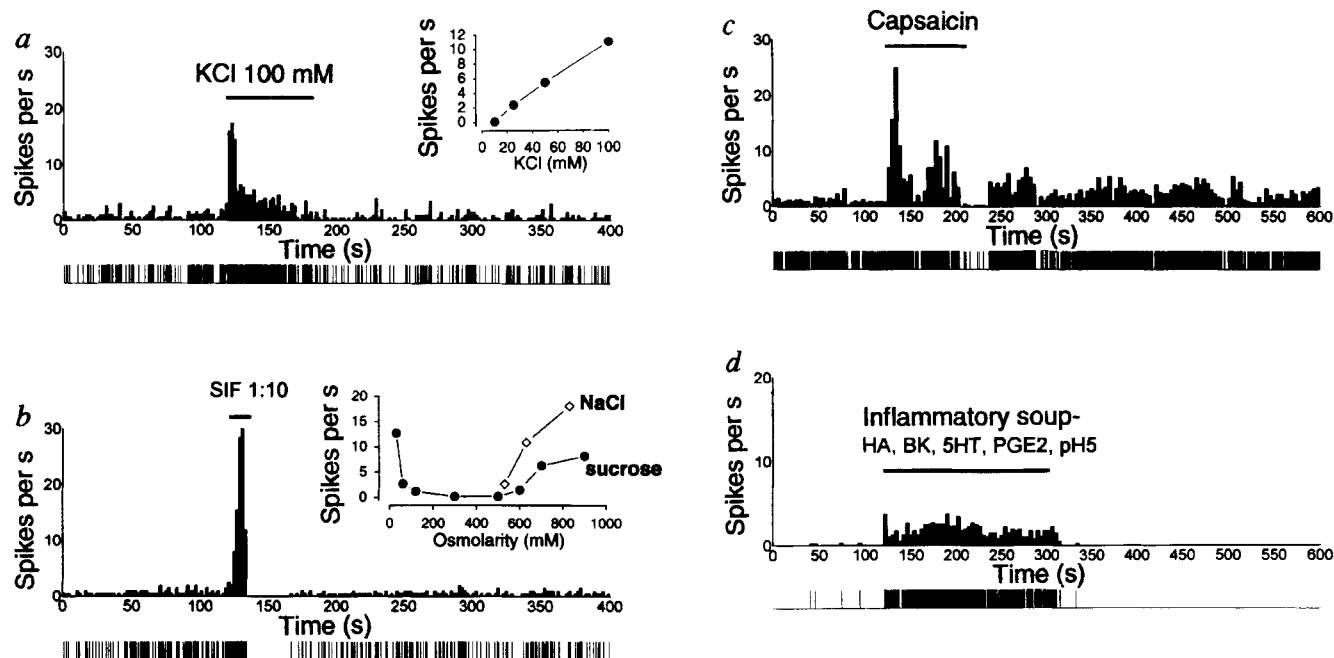


FIG. 2 Responses to topical application to the dura of algesic and inflammatory agents. *a*, Response of a neuron (same cell shown in the right-hand trace of Fig. 1b, c.v. 1.2 ms^{-1}) to 100 mM KCl in SIF, pH 7.2. Inset shows a plot of neuronal discharge against KCl concentration for this neuron. In neurons tested at lower concentrations, responses were found to 25 mM (3/6) and 50 mM KCl (6/6). *b*, Response of the cell shown in *a* to SIF diluted 10-fold with distilled water, pH 7.2. Inset shows plot of neuronal discharge against osmolarity for this neuron. Filled symbols: osmolarity of isotonic SIF was increased by adding increasing concentrations of sucrose (up to 600 mM), and was decreased by diluting the isotonic SIF with distilled water. Open symbols: osmolarity was increased by increasing NaCl concentrations (up to 400 mM) in SIF. The neurons responded to both high and

low osmolarity. Increased NaCl concentrations produced a greater response than sucrose solutions of equal osmolarity in 5/5 neurons, indicating that the response to NaCl is not solely an effect of hyperosmolarity. Responses were found in 5/5 neurons to 250 mM NaCl, which was the lowest NaCl concentration tested. In neurons tested at multiple osmolarities (raised or lowered with sucrose or water, respectively), responses were found to 120 mOsm (5/8), 60 mOsm (6/8), 30 mOsm (8/8), 600 mOsm (1/4), 700 mOsm (3/5), and 900 mOsm (4/5). *c*, Response of a cell (c.v. 3.5 ms^{-1}) to 1% capsaicin. No response was produced by capsaicin vehicle, which was tested immediately before the capsaicin. *d*, Response of a cell (c.v. 4.2 ms^{-1}) to the soup of bradykinin, serotonin, and prostaglandin E_2 , all $10 \mu\text{M}$, and histamine, $100 \mu\text{M}$, SIF, pH 5.0.

into the lumen of the sinus, using hypertonic sodium chloride as a stimulus. Four of these 5 neurons were found to respond to both routes of application (Fig. 1e–g). The response for both routes of chemical application provides evidence that the neurons have receptive endings within or adjacent to the walls of the sinus, consistent with anatomical data¹³.

Sensitization was investigated by determining mechanical sensitivity before and after topical application to the receptive field of low pH buffer ($n = 2$) or the inflammatory soup ($n = 15$), both of which are effective for sensitizing cutaneous nociceptive afferents to mechanical stimuli^{17,18}. Application of these agents for 5 min enhanced the sensitivity to subsequent testing with mechanical stimulation in 10 of 15 mechanosensitive neurons tested, and also resulted in the appearance of mechanosensitivity in 1 of 2 neurons that initially had been unresponsive to our mechanical stimulation. This chemically induced sensitization to mechanical stimuli was significant ($P < 0.01$, Wilcoxon's signed rank test). Such sensitization to mechanical stimuli was observed both in neurons that discharged in response to the sensitizing agent ($n = 6$) as well as in neurons that did not ($n = 5$).

Several previous studies have described responses of central neurons in the spinal trigeminal nucleus and thalamus to electrical, mechanical, and chemical (bradykinin, capsaicin) stimulation of the dura, although sensitization was not examined^{4–12}. These studies showed that central dura-responsive neurons typically had nociceptive facial receptive fields with a distribution reminiscent of the pattern of pain referral evoked by dural stimulation, consistent with a role for these neurons in the mediation of referred pain of intracranial origin. The only prior electrophysiological study of meningeal primary afferents examined responses to mechanical and thermal, but not chemical,

stimuli¹¹. We report here the first direct observations, to our knowledge, of chemosensitivity and chemically induced sensitization to mechanical stimuli in afferents supplying the intracranial meninges, properties that have been found previously for other types of somatic and visceral afferents^{17–20} and that have long been hypothesized in theories of headache pathogenesis^{1,15}, owing to their parallels to the symptomatology of some types of clinically occurring headaches. The headache that accompanies meningitis, as well as migraine headache, is characterized by features that suggest an exaggerated intracranial mechanosensitivity. Normally innocuous activities that produce an increase in intracranial pressure or altered intracranial haemodynamics, such as sudden head movement, coughing, or breath-holding, evoke dramatically increased head pain during such headaches^{1,3}.

Mechanical sensitization might also contribute to the throbbing pain of migraine. Although this pain has long been attributed to the pulsations of abnormally dilated blood vessels^{1,2,21,22}, more recent studies of vessel diameter during migraine attacks have failed to find dilatation greater than that known to occur in the absence of headache^{15,22–24}. It thus appears that cerebral vasodilatation could only contribute to migraine headache if it were accompanied by sensitization of intracranial afferents (or possibly of central neurons^{25,26}) to respond to normally innocuous mechanical stimuli. Our results show that chemically induced sensitization to mechanical stimuli can occur in meningeal afferents. It is generally assumed that the location of the intracranial afferent endings in question must be peri-arterial, because of the throbbing quality of the headache^{1,2,21}. But we note that this need not be the case, because the arterial pressure pulse is transmitted mechanically throughout the subarachnoid space and meninges²⁷.

How might sensitization occur during migraine? Alternatives to

the vasodilatation theory have implicated intracranial processes such as cortical spreading depression²⁸ or meningeal neurogenic inflammation^{15,29}. Cortical spreading depression consists of a slowly propagating depression of neural activity which is accompanied by chemical disturbances, including elevated levels of extracellular potassium²⁸. Meningeal neurogenic inflammation apparently results from release of neuropeptides (for example, substance P, calcitonin-gene-related peptide) from the peripheral endings of dural afferents, which in turn induces degranulation of dural mast cells¹⁵ (whose released contents may include histamine, serotonin and cytokines). Thus, dural sinus afferents may be affected by chemical alterations in the cerebral cortex and the dura itself, because the sinuses collect the venous drainage from

much of the cortex. In addition, recent evidence suggests that substances may move between the dura and underlying cortex by diffusion through the subarachnoid space³⁰.

Our results show that exposure to chemical agents associated with tissue injury or inflammation produces discharge in some meningeal afferents. Such exposure also causes some of these afferents to become more sensitive to mechanical stimuli for up to an hour or more. We suggest that such chemically induced transient mechanical sensitization could lead these neurons to respond to relatively small, normally innocuous mechanical alterations in the intracranial space such as might be produced by haemodynamic changes or head movement, and that excitation of such afferents could contribute to or worsen certain headaches. □

Methods

Urethane-anaesthetized male rats (400–600 g) were placed in a stereotaxic headholder. The right transverse sinus and the caudal part of the superior sagittal sinus were exposed by craniotomy. The exposed dura was bathed in a synthetic interstitial fluid (SIF) consisting of 135 mM NaCl, 5 mM KCl, 1 mM MgCl₂, 5 mM CaCl₂, 10 mM HEPES, and 10 mM glucose, pH 7.2. In 5 animals the entire superior sagittal sinus was exposed, and PE-10 tubing was inserted into the rostral end and passed caudally to the caudal end (sinus confluens). Tungsten microelectrodes were advanced into the trigeminal ganglion by a dorsal approach through the cerebral cortex. Single-shock electrical search stimuli (0.5 ms pulses, 5 mA, 0.7 Hz) were delivered through a bipolar stimulating electrode placed on the dura overlying the ipsilateral transverse sinus while the microelectrode was advanced through the ganglion. Chemical agents were delivered by topical application to the exposed dura or by infusion into the superior sagittal sinus. Sinus infusions were followed by control infusions into the jugular vein. Because the direction of venous blood flow is from the dural sinuses through the jugular vein to the heart (Fig. 1a), the absence of a response to jugular infusion provides evidence that the response to sinus infusion was from a site of action between the sinus and the jugular vein. The jugular infusate would be expected to access the intracranial circulation only after traversing the heart, at which point it is apparently too diluted to produce a response (Fig. 1g). A second sinus infusion was delivered following the jugular infusion, in order to verify that the neuron was still responsive. All agents tested in this study were delivered in SIF, except capsaicin, which was prepared by first dissolving in ethanol, then heating and re-dissolving in 10% Tween-80. Neurons that were tested to more than one agent ($n = 20$) were tested in the following order (following mechanical testing), at 10-min intervals: KCl, NaCl, low osmolarity SIF, high osmolarity SIF, and capsaicin; the inflammatory soup was either tested immediately before the capsaicin or in neurons that were tested for sensitization (Fig. 3), it was the first agent tested. Neuronal responses were quantified by calculating the mean firing rate during the initial 10 s of the response, minus the mean firing rate during the 60 s preceding the stimulus.

To investigate sensitization (Fig. 3), a graded series of von Frey monofilaments (Stoelting, Wood Dale, Illinois) was applied in order of increasing stiffness to measure mechanical threshold at 10–30-min intervals for 1–2 h before and 1–2 h after the conditioning stimulus. Each data point (Fig. 3a) was obtained by applying a single monofilament 5–10 times to the neuron's mechanical receptive field on the ipsilateral transverse sinus. The percentage of times that the neuron responded was plotted against the force of the monofilament. Sensitization was not tested unless the baseline measurements were stable. The plotted post-stimulus measurements (Fig. 3c) were taken 5–30 min after the stimulus, which appeared to be the period of maximum sensitization.

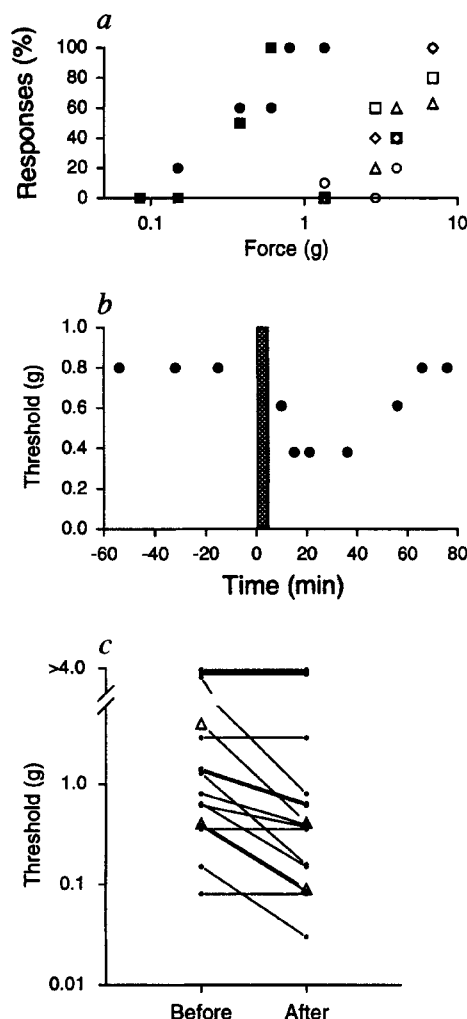


FIG. 3 Mechanical sensitization, as demonstrated by a lowered von Frey threshold, following topical application to the dura of acidic or inflammatory agents. **a**, Decrease in mechanical threshold of a neuron (c.v. 13.9 ms⁻¹) following application of pH 5.0 SIF for 2 min. The open and filled symbols represent responses to von Frey monofilaments before and after the chemical stimulation, respectively. A full series of von Frey responses was tested at 65, 45, 20, and 5 min before the chemical application (open circles, squares, diamonds, and triangles, respectively), and at 5 and 15 minutes following the chemical application (filled circles and squares, respectively). The von Frey threshold (>50% incidence of responses) decreased from about 4.0 g to 0.38 g following exposure to the low pH solution. **b**, Time course of changes in mechanical threshold for another neuron (same cell shown in Fig. 2c) which received a 5-min exposure (starting at time 0) to the soup of bradykinin, serotonin, prostaglandin E₂, and histamine in SIF, pH 5.0. **c**, Plot of mechanical thresholds before and after application of chemical agents. Circles and triangles represent neurons tested with inflammatory soup ($n = 15$) and pH 5.0 buffer ($n = 2$), respectively.

Received 1 August; accepted 23 September 1996.

- Wolff, H. G. *Headache and Other Head Pain* (Oxford Univ. Press, New York, 1963).
- Martin, J. B. in *Harrison's Principles of Internal Medicine* 12th edn (eds Wilson, J. D. et al.) 108–115 (McGraw-Hill, New York, 1991).
- Blau, J. N. & Dexter, S. L. *Cephalalgia* **1**, 143–147 (1981).
- Strassman, A. S., Mason, P., Moskowitz, M. & Maciewicz, R. J. *Brain Res.* **379**, 242–250 (1986).
- Davis, K. D. & Dostrovsky, J. O. *J. Neurophysiol.* **59**, 648–666 (1988).
- Davis, K. D. & Dostrovsky, J. O. *Brain Res.* **446**, 401–406 (1988).
- Davis, K. D. & Dostrovsky, J. O. *Brain Res.* **454**, 89–100 (1988).
- Zagami, A. S. & Lambert, G. A. *Exp. Brain Res.* **81**, 552–566 (1990).
- Zagami, A. S. & Lambert, G. A. *Neurosci. Lett.* **121**, 187–190 (1991).
- Lambert, G. A., Zagami, A. S., Bogduk, N. & Lance, J. W. *Cephalalgia* **11**, 75–85 (1991).
- Dostrovsky, J. O., Davis, K. D. & Kawakita, K. *Can. J. Physiol. Pharmacol.* **69**, 652–658 (1991).
- Kaube, H., Hoskin, K. L. & Hoskin, K. L. & Goadsby, P. J. *Cephalalgia* **12**, 133–136 (1992).
- Penfield, W. & McNaughton, M. *Arch. Neurol. Psychiat.* **44**, 43–75 (1940).
- Andres, K. H., von Düring, M., Muszynski, K. & Schmidt, R. F. *Anat. Embryol.* **175**, 289–301 (1987).
- Moskowitz, M. A. & Macfarlane, R. *Cerebrovasc. Brain Metab. Rev.* **5**, 159–177 (1993).
- Strassman, A. M., Mineta, Y. & Vos, B. P. *J. Neurosci.* **14**, 3725–3735 (1994).
- Steen, K. H., Reeh, P. W., Anton, F. & Handwerker, H. O. *J. Neurosci.* **12**, 86–95 (1992).
- Davis, K. D., Meyer, R. A. & Campbell, J. N. *J. Neurophysiol.* **69**, 1071–1081 (1993).
- Baker, D. G., Coleridge, H. M., Coleridge, J. C. G. & Nerdrum, T. J. *Physiol.* **306**, 519–536 (1980).

20. Schaible, H.-G. & Schmidt, R. F. *J. Physiol. (Lond.)* **403**, 91–104 (1988).
21. Graham, J. R. & Wolff, H. G. *Arch. Neurol. Psychiatr.* **39**, 737–763 (1938).
22. Friberg, L., Olesen, J., Iversen, H. K. & Sperling, B. *Lancet* **338**, 13–17 (1991).
23. Iversen, H. K., Nielsen, T. H., Olesen, J. & Tfelt-Hansen, P. *Lancet* **336**, 837–839 (1991).
24. Thomsen, L. L., Iversen, H. K. & Olesen, J. *Cephalalgia* **15**, 109–116 (1995).
25. Woolf, C. J. *Nature* **308**, 686–688 (1983).
26. Burstein, R., Malick, A. & Strassman, A. *Soc. Neurosci. Abstr.* **22**, 864 (1996).
27. Daley, M. L., Pasupathy, H., Griffith, M., Robertson, J. T. & Leffler, C. W. *IEEE Trans. Biomed. Eng.* **42**, 420–424 (1995).
28. Lauritzen, M. *Brain* **117**, 199–210 (1994).
29. Goadsby, P. J. & Edvinsson, L. *Ann. Neurol.* **33**, 48–56 (1993).
30. Orr, E. L. & Stokely, M. E. *Soc. Neurosci. Abstr.* **21**, 1858 (1995).

CORRESPONDENCE and requests for materials should be addressed to A.M.S. (e-mail: andrew.strassman@biol.harvard.edu).

From stimulus encoding to feature extraction in weakly electric fish

F. Gabbiani*, W. Metzner†, R. Wessel‡ & C. Koch*

* Computation and Neural Systems Program, 139-74 Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

† Department of Biology, University of California at Riverside, Riverside, California 92521, USA

‡ Department of Biology, University of California at San Diego, La Jolla, California 92093, USA

ANIMALS acquire information about sensory stimuli around them and encode it using an analogue or a pulse-based code. Behaviourally relevant features need to be extracted from this representation for further processing. In the electrosensory system of weakly electric fish, single P-type electroreceptor afferents accurately encode the time course of random modulations in electric-field amplitude¹. We applied a stimulus estimation method² and a signal-detection method to both P-receptor afferents and their targets, the pyramidal cells in the electrosensory lateral-line lobe. We found that although pyramidal cells do not accurately convey detailed information about the time course of the stimulus, they reliably encode up- and downstrokes of random modulations in electric-field amplitude. The presence of such temporal features is best signalled by short bursts of spikes, probably caused by dendritic processing, rather than by isolated spikes. Furthermore, pyramidal cells outperform P-receptor afferents in signalling the presence of temporal features in the stimulus waveform. We conclude that the sensory neurons are specialized to acquire information accurately with little processing, whereas the following stage extracts behaviourally relevant features, thus performing a nonlinear pattern-recognition task.

The electric fish *Eigenmannia* generates a quasi-sinusoidal electric field with carrier frequencies between 200 and 600 Hz, by regularly discharging an electric organ located in its tail. The fish senses local distortions of the electric field by means of two classes of tuberous electroreceptors distributed on the body surface³: T-type and P-type electroreceptors encode changes in phase and electric field amplitude, respectively, which are used for electrolocation⁴ and communication⁵.

At the level of the electrosensory lateral line-lobe (ELL), the first central nucleus of the electrosensory pathway, amplitude information is processed nearly independently of phase information, and is re-encoded in the output spike trains of E- and I-type pyramidal cells^{4,6}. The temporal response of P-receptor afferents and pyramidal cells has been characterized by their average firing rate for step changes and sinusoidal amplitude modulations of an externally applied electric field^{4,7}. Under these conditions, the response of P-receptor afferents is slowly adapting (tonic), whereas the response of pyramidal cells is transient (phasic).

Furthermore, P-receptor afferents and E-type pyramidal cells raise their mean firing rate when the electric field amplitude is increased, whereas I-type pyramidal cells are inhibited. These mean response characteristics leave open several alternatives for the encoding and processing of time-varying modulations in electric-field amplitude in single-spike trains of ELL pyramidal cells. Pyramidal cells might, for instance, transmit detailed information about the time course of temporal changes in the stimulus waveform⁸, or of specific stimulus frequencies⁹, combined with half-wave rectification.

To address the question of how modulations in electric-field amplitude are temporally encoded and processed between the first two stages of the amplitude pathway, we studied the responses of P-receptor afferents and pyramidal cells to random distortions of a mimic of the fish's own electric field. Such distortions were simulated by superposing zero-mean, random amplitude modulations, $s(t)$, to an externally applied electric field in fish whose electric-organ discharges were strongly attenuated by intramuscular injection of a curare-like drug (Fig. 1a). The carrier frequency of this field was equal to the fish's own frequency before the drug was applied, and the cut-off frequency of the random amplitude modulations¹ (Fig. 1c) was varied over a behaviourally relevant range (2–40 Hz) (refs 4, 10).

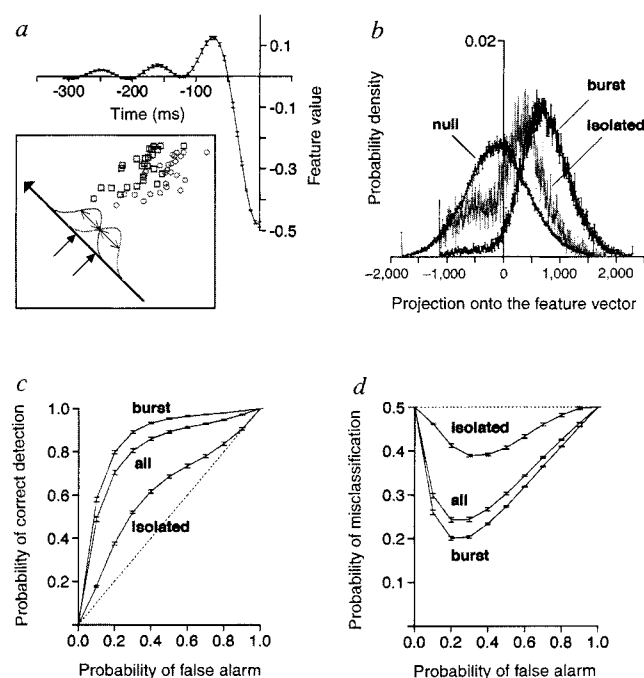
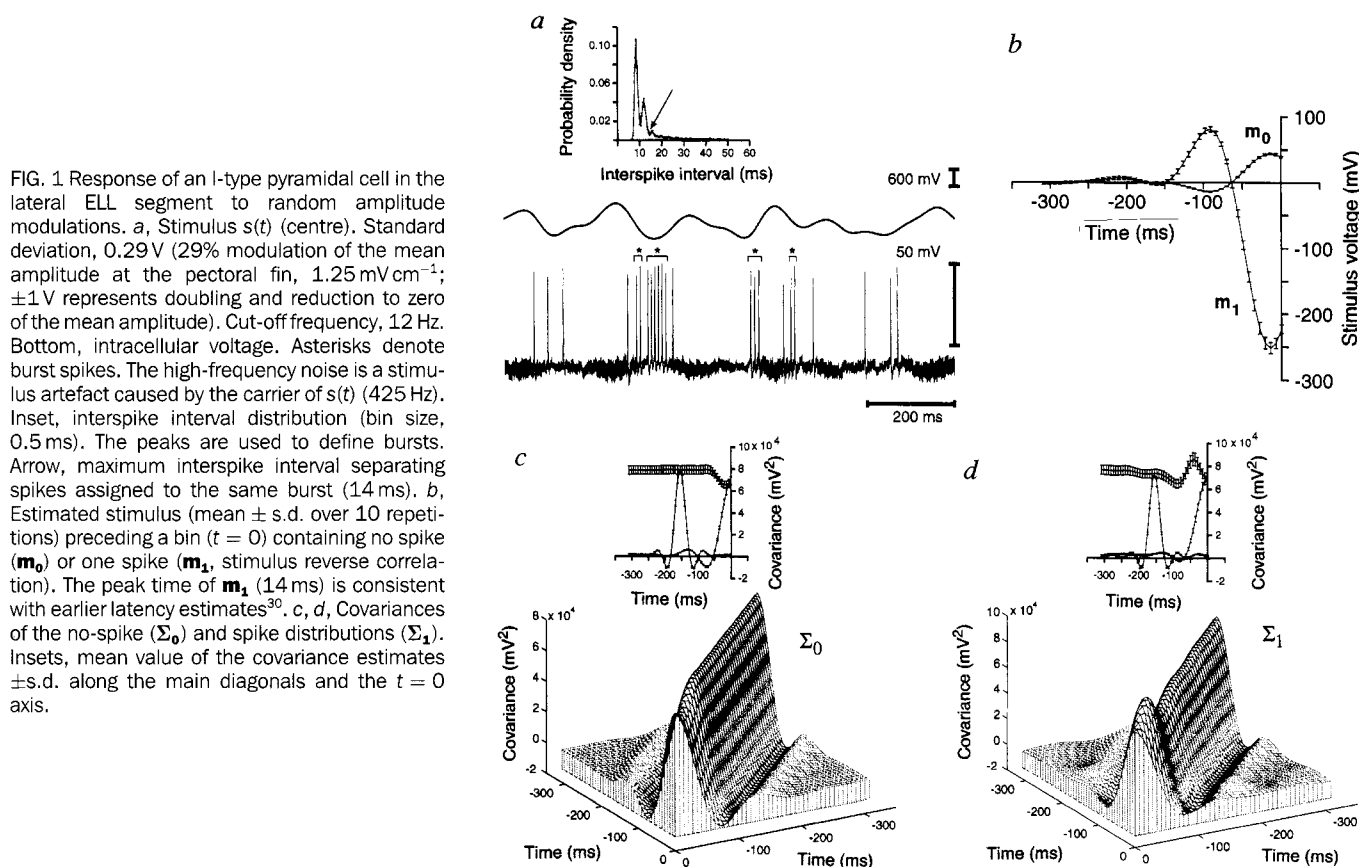
We assessed the ability of single-spike trains to convey detailed information about the time course of $s(t)$ by smoothing the time series of action potential events with a filter chosen to obtain the best estimate of $s(t)$ (refs 1, 2, 11). This method and its variants are able to detect detailed encoding of a random stimulus or some filtered and/or half-wave rectified version of it (such as temporal derivative encoding, band-pass filtering and on/off responses). The accuracy of the information transmitted about $s(t)$ was characterized in the time domain by the coding fraction, defined as $\gamma = 1 - \varepsilon/\sigma$, where ε is the root-mean-square error between the true and estimated stimulus, and σ is the standard deviation of $s(t)$ (ref. 1). The coding fraction takes the maximum value of 1 when the stimulus is perfectly estimated ($\varepsilon = 0$) and the minimum value of 0 if estimation from the spike train is at chance level ($\varepsilon = \sigma$) (refs 12, 13). We computed the coding fraction for P-receptor afferents and pyramidal cells; single P-receptor afferents were able to encode up to 75% of the stimulus time course, but pyramidal cells encoded less than 30%. Such a poor performance suggests that individual pyramidal cells do not convey detailed information on the time course of the stimulus to higher-order levels of the electrosensory system.

Pyramidal cells might transmit information about temporal features in the stimulus by performing a nonlinear classification task. This possibility can be tested by using neural network models. An artificial neural network can learn from the stimulus and response data to find the 'optimal' sensory input pattern eliciting a spike¹⁴. We used a more direct method derived from signal-detection theory, which assumes that the nonlinear classification task is implemented by means of a linear summation of incoming information plus a threshold computation^{11,15}. To determine the optimal temporal feature $\mathbf{f}$ predicting the occurrence or non-occurrence of a spike in a pyramidal cell, we binned the data and studied the distributions of amplitude modulation waveforms preceding a bin containing a spike or no spike. We first determined the mean stimulus preceding a bin containing a spike ($\mathbf{m}_1$) and no spike ($\mathbf{m}_0$) (Fig. 1b) and then computed the covariance matrices Σ_1 and Σ_0 (Fig. 1c, d) characterizing second-order variations and correlations of these distributions around their means. The optimal $\mathbf{f}$ was determined as the direction in this temporal stimulus space of amplitude modulation waveforms for which the ratio of the mean squared distance between stimuli preceding a spike and no spike (computed from $\mathbf{m}_1$ and $\mathbf{m}_0$) and of their variances (computed from Σ_1 and Σ_0) was maximized (see equation (1) in Methods). This method is illustrated in a two-dimensional example (Fig. 2a, inset). For the I-type pyramidal cell depicted in Figs 1 and 2, the optimal sensory stimulus that predicted the occurrence of a spike, $\mathbf{f}$, was a downstroke in the

electric-field amplitude, preceded at around 100 ms by a small upstroke (Fig. 2a). For E-type pyramidal cells, the shape of the optimal feature was reversed. The time course of f (Fig. 2a) differs from the time course of the mean stimulus preceding a spike, m_1 (Fig. 1b), because the former takes into account the mean stimulus preceding a bin containing no spike, m_0 , as well as second-order variations, Σ_1 and Σ_0 . In general f did not differ significantly from

zero for time values of 300 ms or more before the spike occurrence (Fig. 2a).

A large fraction ($56 \pm 21\%$, $n = 30$ cells) of the spikes generated by pyramidal cells in response to random amplitude modulations occurred in short bursts (mean number of spikes per burst, 2.9 ± 1.3 ; Fig. 1a, inset). To examine whether the occurrence of burst-like spike patterns signalled the presence of the temporal



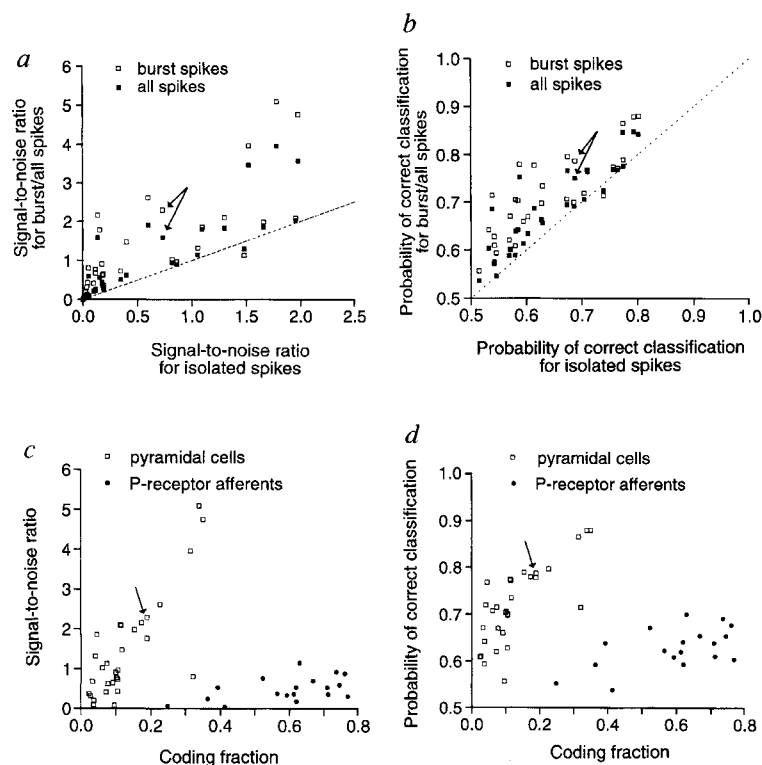


FIG. 3 Pyramidal cells ($n = 30$) extract temporal features from the stimulus time course encoded by P-receptor afferents ($n = 18$). *a*, Signal-to-noise ratios in the discrimination task (see equation (1)) for isolated spikes plotted against signal-to-noise ratio for burst spikes (open squares) and all spikes (filled squares). Dashed line, identical performance. Arrows, data points obtained from the cell used in Figs 1 and 2. *b*, Probability of correct classification for isolated spikes ($1 - \text{minimum probability of misclassification}$; see Fig. 2*d*) plotted against the probability of correct classification for burst spikes (open squares) and all spikes (filled squares). *c*, Signal-to-noise ratio as a function of coding fraction for pyramidal cells (open squares, burst spikes) and P-receptor afferents (filled circles). Note the weak and moderate correlation between these variables for P-receptor afferents and pyramidal cells (Pearson correlation coefficient, $\rho = 0.50$ and 0.80). For each pyramidal cell and P-receptor afferent the best performance in the discrimination task was selected. *d*, Probability of correct classification as a function of coding fraction (as in *c*; $\rho = 0.58$ and 0.76). Note the larger variability in the performance of pyramidal cells.

feature vector more reliably than isolated spikes, we separated the stimuli preceding a spike into two groups, depending on whether the spike was part of a burst or not, before computing their projection onto the feature vector (Fig. 2*b*). The distribution of stimuli associated with burst spikes had a significantly smaller overlap with the distribution of stimuli preceding no-spike containing bins (the null distribution in Fig. 2*b*) than the distribution of stimuli associated with isolated spikes. This was confirmed by comparing the receiver-operating characteristics curves describing the probability that an ideal observer will correctly identify a stimulus as eliciting a spike (probability of correct detection) for a fixed probability of incorrectly identifying a stimulus as eliciting a spike (probability of false alarm). For any false alarm rate, spikes from bursts were a better indicator for the presence of the optimal sensory stimulus than isolated spikes or all spikes taken together (Fig. 2*c*). This was also true when considering the probability of misclassification of a stimulus, given by averaging the probability of incorrectly identifying a stimulus as eliciting a spike (probability of false alarm) and the probability of incorrectly identifying a stimulus as eliciting no spike (one minus the probability of correct detection; Fig. 2*d*). In this particular example, the performance of all spikes taken together was close to the performance of spikes occurring in bursts, reflecting the fact that most spikes (77%) did occur in bursts. Similar results were found for all 30 pyramidal cells analysed and are summarized in Fig. 3*a, b*.

How accurately do processes at the input stage to the ELL (P-receptor afferents) convey explicit information about the presence or absence of temporal features? After repeating the same analysis on spikes of P-receptor afferents, we found that our sample population of P-receptor afferents ($n = 18$) was significantly outperformed in the discrimination task. The signal-to-noise ratio achieved by pyramidal cells in the discrimination task reached maximal values of 4.0 for all spikes combined, and up to 5.0 when only spikes belonging to bursts were considered (Fig. 3*a, c*). Accordingly, by integrating their inputs over the previous 300 ms and generating a spike, pyramidal cells could indicate the presence of the optimal feature with 85% accuracy, and up to 88% accuracy when only spikes belonging to bursts are considered (the chance level is 50%; Fig. 3*b, d*). But, for P-receptor afferents, the maximal signal-to-noise ratio in the discrimination task was

only 1.2 and the discrimination accuracy 70% (Fig. 3*c, d*). The difference between the discrimination ability of all pyramidal cell spikes and of P-receptor afferent spikes was statistically significant (Wilcoxon rank-sum test, $\alpha \leq 0.025$), and the difference between pyramidal cell spikes belonging to bursts and P-receptor afferent spikes was highly significant (Fig. 3*d*, vertical axis; $\alpha \leq 0.0005$). In contrast, the fraction of the stimulus time course encoded by P-receptor afferents was always considerably higher than that encoded by pyramidal cells (Fig. 3*c, d*, horizontal axis). Thus the temporal features are extracted by the ELL circuitry from the detailed time-course information conveyed in the spike trains of P-receptor afferents.

We assumed that feature extraction resulted from a linear summation of incoming temporal information followed by a threshold computation. This computation could be performed at or near the soma of pyramidal cells using the threshold of the spike-generation mechanism. Such a simple model requires information on the detailed time course of the stimulus to be present in the somatic subthreshold membrane potential. To test this, we estimated the stimulus directly from the intracellular subthreshold membrane potential, obtained by removing the spikes from the membrane voltage trace. We then compared the coding fraction with that obtained from pyramidal cell spike events. The fraction of the stimulus encoded in the intracellular membrane potential was not significantly higher than that encoded in spikes (mean increase, $1.7 \pm 1.8\%$; $n = 20$ cells). It is therefore unlikely that temporal feature extraction was computed by means of a simple thresholding of the membrane voltage. In an *in vitro* preparation of the ELL from a closely related species, it has been demonstrated that retrograde propagation of somatic action potentials in conjunction with tetrodotoxin-sensitive conductances localized in the dendrites were responsible for short spike bursts similar to those observed here *in vivo*¹⁶. This observations, combined with our intracellular recordings, strongly suggests that feature extraction results from dendritic processing. Inhibitory control of dendritic pyramidal cell excitability^{4,17} is likely to play an important role in this computation.

ELL pyramidal cells of weakly electric fish exemplify how active dendritic processing can solve a specific temporal pattern-recognition task¹⁸. The short bursts of spikes emitted in response to up-

and downstrokes of electric-field amplitude modulations represent code symbols¹⁹ that can be expected to be robust against subsequent noise sources, such as those introduced by unreliability in synaptic transmission²⁰. This robustness, which might explain the widespread occurrence of burst spike sequences in various sensory systems, is at the expense of temporal resolution. However, the resolution that can be inferred from the average burst duration (18 ± 9 ms first-to-last spikes in 30 cells) is sufficient for the detection of distortions of the electric field during electrolocation⁴ and for the initiation of the jamming avoidance response¹⁰. In contrast, a substantially higher temporal resolution is achieved in the separate channel formed by T-type electroreceptor afferents and higher-order neurons distinct from those processing amplitude modulations²¹. The statistical analysis used here, complementary to the successful neuroethological approach¹⁰, indicates that the electrosensory system in weakly electric fish may help us to understand better the links between biophysics and specific neural computations, such as temporal pattern recognition. □

Methods

Electrophysiology and stimuli. Extracellular recordings from receptor afferents and intracellular recordings from pyramidal cells were performed using standard electrophysiological techniques^{1,22}. Intracellular recording electrodes were filled with neurotracer (neurobiotin; Vector) for intracellular labelling. This allowed to verify the cell type and its location within the ELL. Where no intracellular cell labelling could be performed, the recording site was verified by setting electrolytic lesions at the conclusion of the experiment using a high-frequency current source²³. Standard histological methods were used and sections were analysed using light microscopy following the common nomenclature²⁴. Electrode placements in the central region of the pyramidal cell layer of the ELL are expected to record activity of superficial and intermediate pyramidal cells²⁵. Deep pyramidal cells were not included in our sample. A detailed description of the random stimuli used is given in ref. 1 (see equation (2)).

Data analysis. Data from each E- and I-type pyramidal cell from two different ELL segments (centromedial and lateral²⁶) and from each P-receptor afferent were analysed to assess their ability to convey detailed information about the stimulus time course or the presence of temporal features in the stimulus waveform. Methods of stimulus estimation are described in refs 1, 12. For stimulus estimation from the subthreshold membrane potential, spikes were removed by two methods with similar results: median filtering of the membrane voltage trace; and threshold detection of spikes and linear interpolation of the voltage between the end points of a spike. To assess the ability to convey feature information, spike trains and stimuli of a duration of 135 s were binned using three bin sizes between $\Delta t_{\min} = 0.5$ ms (the sampling rate) and $\Delta t_{\max} =$ maximal bin size, resulting in no more than one spike per bin. Of those, the size yielding best performance was retained²⁷ (in Figs 1 and 2, $\Delta t = 7$ ms). Maximum-likelihood estimators of the mean vectors (respectively of the covariances; ref. 15, section 3.2) had at most 100 (respectively 100×100) time components. The optimal feature $\mathbf{f}$ was obtained by maximizing the signal-to-noise ratio (SNR) (Fisher's linear discriminant function; see ref. 15, section 6.5; ref. 11, section IIIB) defined by,

$$\text{SNR}(\mathbf{f}) = \frac{[\mathbf{f} \cdot (\mathbf{m}_1 - \mathbf{m}_0)]^2}{\mathbf{f} \cdot \frac{1}{2}(\Sigma_0 + \Sigma_1)\mathbf{f}} \quad (1)$$

The vector $\mathbf{f}$ that maximizes SNR($\mathbf{f}$) satisfies $(\mathbf{m}_1 - \mathbf{m}_0) = 1/2(\Sigma_0 + \Sigma_1)\mathbf{f}$. This equation was solved numerically by diagonalizing $\Sigma_0 + \Sigma_1$ using MATLAB subroutines (that is, by performing a principal component analysis) and retaining the first n largest eigenvalues accounting for 99% of the variance (ref. 28, sections 6.1 and 8.1). The receiver-operating characteristics and the minimum probability of error were computed using the resubstitution method²⁹. Typical sample sizes: null distribution, $N > 10,000$; isolated spikes, $N > 500$; burst spikes, $N > 2,000$.

Received 18 June; accepted 11 October 1996.

- Wessel, R., Koch, C. & Gabbiani, F. *J. Neurophysiol.* **75**, 2280–2293 (1996).
- Bialek, W., de Ruyter van Steveninck, R. & Wariand, D. *Science* **252**, 1854–1857 (1991).
- Zakon, H. in *Electroreception* (eds Bullock, T. H. & Heiligenberg, W.) 103–156 (Wiley, New York, 1986).
- Bastian, J. in *Electroreception* (eds Bullock, T. H. & Heiligenberg, W.) 577–611 (Wiley, New York, 1986).
- Hopkins, C. *Annu. Rev. Neurosci.* **11**, 497–535 (1988).
- Maler, L., Sas, E. & Rogers, J. *J. Comp. Neurol.* **195**, 87–139 (1981).
- Hopkins, C. *J. Comp. Neurol.* **111**, 171–207 (1976).
- Bastian, J. & Heiligenberg, W. *J. Comp. Neurol.* **136**, 135–152 (1980).
- Turner, R., Plant, J. & Maler, L. *J. Neurophysiol.* **76**, 2364–2382 (1996).
- Heiligenberg, W. *Neural Nets in Electric Fish* (MIT Press, Cambridge, MA, 1991).
- Poor, H. *An Introduction to Signal Detection and Estimation* (Springer, New York, 1994).
- Gabbiani, F. & Koch, C. *Neural Computat.* **8**, 44–66 (1996).
- Gabbiani, F. *Network: Comp. Neural Syst.* **7**, 61–85 (1996).

- Lehky, S., Sejnowski, T. & Desimone, R. *J. Neurosci.* **12**, 3568–3581 (1992).
- Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 1984).
- Turner, R. W., Maler, L., Deerinck, T., Levinson, S. R. & Ellisman, M. H. *J. Neurosci.* **14**, 6453–6471 (1994).
- Maler, L. & Mugnaini, E. *J. Comp. Neurol.* **345**, 224–252 (1994).
- Yuste, R. & Tank, W. *Neuron* **16**, 701–716 (1996).
- Bair, W., Koch, C., Newsome, W. & Britten, K. *J. Neurosci.* **14**, 2870–2892 (1994).
- Allen, C. & Stevens, C. *Proc. Natl Acad. Sci. USA* **91**, 10380–10383 (1994).
- Kawasaki, M., Rose, G. & Heiligenberg, W. *Nature* **336**, 173–176 (1988).
- Metzner, W. & Heiligenberg, W. *J. Comp. Neurol.* **169**, 135–150 (1991).
- Metzner, W. *J. Neurosci.* **13**, 1862–1878 (1993).
- Maler, L., Sas, E., Johnston, S. & Ellis, W. *J. Chem. Neuroanat.* **4**, 1–38 (1991).
- Bastian, J. & Courtright, J. *J. Comp. Neurol.* **168**, 393–397 (1991).
- Carr, C., Maler, L. & Sas, E. *J. Comp. Neurol.* **211**, 139–153 (1982).
- Krueger, J. & Becker, J. *Trends Neurosci.* **14**, 282–286 (1991).
- Jolliffe, I. *Principal Component Analysis* (Springer, New York, 1986).
- Raudys, S. & Jain, A. *IEEE Trans. Patt. Anal. Mach. Intell.* **13**, 252–264 (1991).
- Shumway, C. *J. Neurosci.* **9**, 4388–4399 (1989).

ACKNOWLEDGEMENTS. We thank J. Juraneck for computer assistance, and M. Konishi, W. Kristan and G. Laurent for comments. This work was supported by grants from NSF, NIMH, UCR, the Center for Neuromorphic Systems Engineering as a part of NSF's Engineering Research Center program and the Sloan Center for Theoretical Neuroscience.

CORRESPONDENCE and requests for materials should be addressed to F.G. (e-mail: gabbiani@klab.caltech.edu).

Mutations in the kinase Rsk-2 associated with Coffin–Lowry syndrome

Elisabeth Trivier, Dario De Cesare, Sylvie Jacquot, Solange Pannetier, Elaine Zackai*, Ian Young†, Jean-Louis Mandel, Paolo Sassone-Corsi & André Hanauer

Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS, INSERM, ULP, B.P. 163, 67404 Illkirch, France

* Clinical Genetic Center, The Children's Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA

† Centre for Medical Genetics, City Hospital, Nottingham NG5 1PB, UK

THE Coffin–Lowry syndrome (CLS), an X-linked disorder, is characterized by severe psychomotor retardation, facial and digital dysmorphisms, and progressive skeletal deformations¹. Genetic linkage analysis mapped the CLS locus to an interval of 2–3 megabases at Xp22.2. The gene coding for Rsk-2, a member of the growth-factor-regulated protein kinases, maps within the candidate interval, and was tested as a candidate gene for CLS. Initial screening for mutations in the gene for Rsk-2 in 76 unrelated CLS patients revealed one intragenic deletion, a nonsense, two splice site, and two missense mutations. The two missenses affect sites critical for the function of Rsk-2. The mutated Rsk-2 proteins were found to be inactive in a S6 kinase assay. These findings provide direct evidence that abnormalities in the MAPK/RSK signalling pathway cause Coffin–Lowry syndrome.

Genetic mapping studies place the CLS locus in Xp22.2, within an interval of approximately 3 cM, between DXS365 and DXS7161 (refs 2–4). During a search for expressed genes, the sequence of a subclone derived from a yeast artificial chromosome insert (YAC ywxd2527), included in a contig encompassing the CLS interval (E.T. *et al.*, manuscript in preparation), showed 100% homology with a complementary DNA coding for Rsk-2 (ref. 5); this localization has been confirmed independently⁶.

Rsk-2 is a member of a family of growth-factor-regulated serine–threonine kinases, known as p90^{sk} or RSK (for ribosomal S6 kinase). Homologues of RSK exist in several species^{3,7–11}. In humans the RSK family comprises three closely related members: Rsk-1, Rsk-2 and Rsk-3 (refs 10, 11). The cDNA that encodes Rsk-2 has an open reading frame of 2,220 bases, encoding a protein of 740 amino acids⁵. A highly conserved feature of all members of the RSK family is the presence of two non-identical, kinase catalytic domains¹². RSKs

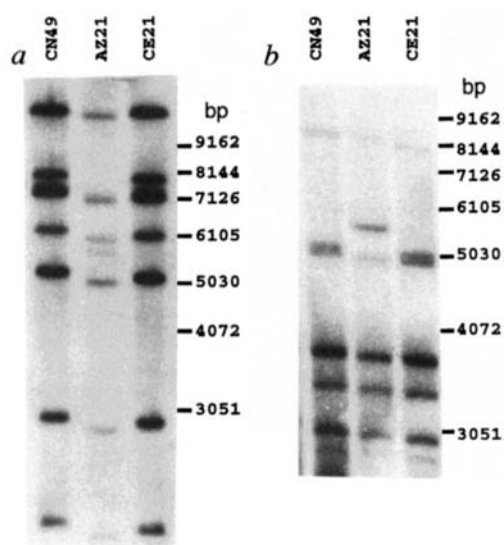


FIG. 1 Identification of an intragenic deletion in a CLS patient. *a*, Genomic DNA from 76 CLS patients and one normal control individual was digested with *Eco*RI and hybridized with a complete *Rsk-2* cDNA probe. Representative results from three unrelated patients are shown. The cDNA detected a 5.8-kb *Eco*RI fragment, instead of the normal 7.8-kb fragment, in patient AZ21. All other patients show the same pattern as the normal control. *b*, Probing of a filter containing *Taq*I-digested genomic DNA with a cDNA probe encompassing nucleotides 300–1,100 also revealed an abnormal *Taq*I fragment in patient AZ21. Patterns of all other patients were similar to that of the normal control. The presence of an intragenic deletion, including 187 bp of coding sequence, in patient AZ21 was confirmed by RT–PCR.

are implicated in the activation of the mitogen-activated kinase cascade and the stimulation of cell proliferation (at the transition between phases G0 and G1 of the cell cycle) and differentiation¹³.

The localization of the gene for *Rsk-2* within the CLS interval, together with its role in signalling pathways, prompted us to investigate a possible involvement in CLS. A probe corresponding to the *Rsk-2* coding region was hybridized to Southern blots of patient DNAs for the detection of fragments of unusual size or the loss of fragments. Patient samples from 76 families were screened, and one patient (AZ21) had a genomic deletion of ~2 kilobases (Fig. 1). Amplification by polymerase chain reaction after reverse transcription of cDNA (RT-PCR) from the patient and direct sequencing showed a deletion of 187 base pairs between positions 406 and 593. The deletion produces a frameshift, generating a TAA termination codon 33 bp downstream of the deletion junction (Fig. 2). The mutation cosegregated with CLS in two affected males and one female with discrete manifestations in this family (data not shown).

We then searched for point mutations. As the intron-exon structure of the gene for *Rsk-2* was not known, 20 pairs of primers were designed throughout the coding region to test for amplification of genomic DNA. In three instances a product of expected size was generated, which then could be subjected to single-strand conformation polymorphism (SSCP) analysis. PCR products from 76 patients, obtained with the three primer sets and encompassing 310 bp of the coding sequence, were analysed using SSCP. Sequencing of one variant band revealed a C to T transition at nucleotide 1,672 in patient AH72, which converts an arginine to a TGA stop codon (Fig. 2). It was transmitted from his mother to the proband but not to three healthy sons or one non-manifesting daughter (Fig. 3a). Analysis of RT-PCR products from probands of five kindreds, of which we established lymphoblastoid or fibroblast cell lines, revealed four additional mutations. A smaller product was observed in patient AB21; the region between nucleotides 1,781 and 2,028 was amplified (Fig. 3b). Sequencing of the product

FIG. 2 Rsk-2 mutations in Coffin-Lowry syndrome patients. Capital letters indicate coding sequences; lower-case letters indicate intronic sequences. Top numbers refer to nucleotide positions, bottom numbers to amino-acid positions. The nucleotide sequence differences between the wild-type alleles and the mutants alleles are indicated in bold letters in the second column. Predicted differences in the polypeptide sequence are indicated in the third column. Right, schematic representations of the predicted Rsk-2 transcripts. ATP binding sites are shown as black boxes, and APE sites as hatched boxes. The presence of an aberrant COOH-terminal peptide in patient DX34 is indicated by a grey box. The position of the missense mutations found in patients AG96 and CN14 are indicated by asterisks.

Patient	Mutation	Predicted protein
AZ21	399 594 CATTATG.//.TAT DEL	CATTATG TAT ACTTCTTGGATGAAGAGGTCACATCAAGTTAA H Y V Y F L M K K V T S S * 146
WT		399 594 CATTATG.....TATACTTCTTGGATGAA H Y A.....I L L D E 134 199 740
AH72	TCTATT TGA AATT	TCTGGTAATCCGGAATCTATT TGA S G N P E S I * 557
WT	1672 TCTATT CGA AATT	1651 TCTGGTAATCCGGAATCTATT CGA ATTGATTTT S G N P E S I R I D F 551
AB21	aattaa a TTACAC	ATGCTTACCGG GAC CTTGGTGTCAAAAGATGCTTCATG TAG M L T G T W C Q R C F M * 622
WT	1842 aattag TTA CAC	1831 1960 ATGCTTACCGG.....GACCTGGTGTCT M L T G.....D L V S 611 654
DX34	GTTCAG gtt tca	1222 GTTCAG gtt tcaaaacttctcatttgc aa aggta ctt gaaaac ct atag V Q A S N F S F F A K V L E N L * 408 422
WT	1222 GTTCAG gtt tca	
CN14	TTAGTGCAGGG A	AAAGTATTAT GTG CAGGGATCATTTGGAAAGGTTTCTTA K V L V Q G S F G K V F L 740
WT	224 TTAGGGCAGGG A	214 AAAGTATTAT GGG CAGGGATCATTTGGAAAGGTTTCTTA K V L G Q G S F G K V F L 72
AG96	GCATAT GCT TTT	GAAAAGAAGGCATAT GCT TTTGTGGA ACT GTG E K K A Y A F C G T V 740
WT	679 GCATAT TC TTT	664 GAAAAGAAGGCATAT TC TTTGTGGA ACT GTG E K K A Y S F C G T V 222

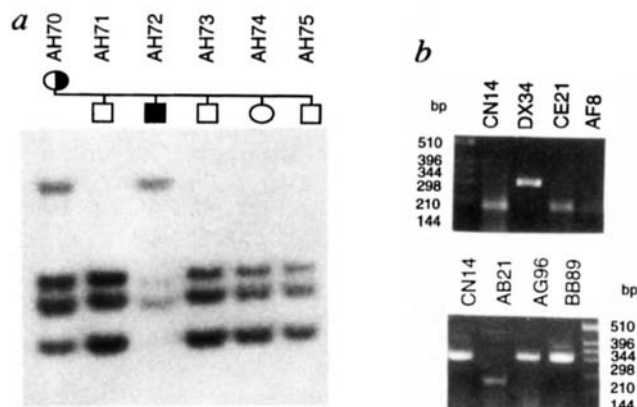


FIG. 3 Mutation analysis. *a*, SSCP analysis in a family with a R558Stop mutation. PCR products encompass bases 1,601–1,763. Proband AH72 shows a band shift caused by a C to T transition at nucleotide 1,672; the unaffected siblings all exhibit a normal band pattern. The mother AH70 has both the shifted and normal bands, confirming that she is CLS carrier. *b*, Detection of splice site mutations in CLS patients. Top, PCR products encompass bases 1,121–1,305. A larger fragment is observed in patient DX34, compared to the normal-size fragment noted in other patients or control subjects. Bottom, PCR products encompass bases 1,781–2,028. An RT-PCR product smaller than was found in other patients or in normal control subjects is observed in patient AB21. In each panel, representative results from three unrelated patients are shown. RT-PCR products were resolved by electrophoresis in a 2% agarose gel.

revealed a deletion of 118 bp, beginning at nucleotide 1,842. This results in a frameshift and premature termination (Fig. 2).

When the region between nucleotides 1,121 and 1,305 was amplified in patient DX34, a larger product was obtained (Fig. 3b). Direct sequencing revealed an insertion of 100 bp at nucleotide 1,227. This insertion codes for an aberrant sequence of 13 amino acids followed by a stop codon. We determined the exon–intron boundaries in the corresponding regions of the gene by PCR. Sequencing revealed a transition of G to A of a splice acceptor site (position –1) in patient AB21, resulting in aberrant splicing of the exon located immediately downstream (nucleotides 1,842 to 1,859 in the cDNA) (Fig. 2). A donor splice site mutation (GT to GC) was found in patient DX34, resulting in the use of a cryptic donor site 100 bp downstream (Fig. 2).

RT-PCR products from patients CN14 and AG96 showed evidence for sequence variation when analysed by SSCP. Direct sequencing revealed a G to T transversion at base 224 in patient CN14 (glycine to valine substitution at position 75), and a T to G transversion at base 679 in patient AG96 (serine to alanine change at codon 227). Gly 75 is a conserved residue located within the putative ATP-binding site¹⁴. Ser 227 is also conserved, and is believed to be a phosphorylation site of the kinase domain of the N terminus, which is essential for catalytic function^{5,11,15}.

Levels of RSK protein were analysed in the lymphoblastoid cell

lines from patients. We prepared total cell extracts and used antibodies raised against Rsk-1, Rsk-2 and Rsk-3 in western blots. Cells from patients express Rsk-1 and Rsk-3 at levels comparable to normal cells (Fig. 4a). This demonstrates that mutations in the gene coding for Rsk-2, as present in patients, do not result in aberrant expression of Rsk-1 and Rsk-3. We also show that the AG96 and CN14 mutated proteins are expressed at similar levels to the wild-type Rsk-2 (Fig. 4a). The truncated Rsk-2 from the AB21 and AZ21 lines are not detected because the epitope of the anti-Rsk-2 antibody corresponds to the deleted C-terminal part of the protein.

We then tested whether the mutations in the gene for Rsk-2 causing CLS result in proteins with impaired biological activity. Phosphorylation activity of the S6 ribosomal substrate was tested on extracts immunoprecipitated with the anti-Rsk-2 antibodies. S6 is readily phosphorylated by the wild-type Rsk-2 immunoprecipitated from normal lymphoblastoid cells, but Rsk-2 from patient AG96 has a strikingly impaired activity, which is comparable to the activity of extracts containing no kinase (Fig. 4b). These results demonstrate that the CLS phenotype is directly linked to the loss of the Rsk-2 biological function.

Both the missense mutations G75V (patient CN14) and S227A (patient AG96) affect the amino-catalytic domain. Site-directed mutagenesis of Ser 218 on Rsk-3, homologous to Ser 227 in Rsk-2, showed that it represents a site of autophosphorylation important for maximal enzyme activity¹⁶; our results on Ser 227 in Rsk-2 confirm this observation (Fig. 4b).

Both kinase domains are required for maximal RSK activity in response to growth-factor stimulation^{16,17}. The natural mutations in CLS patients constitute valuable tools for the study of RSK signalling and control. The various nature of the mutations described here, and the fact that over 50% of the families collected are sporadic cases, may indicate the presence of different mutations with variable pathological consequences. Indeed, variable degree in mental retardation has been documented for CLS patients (E.Z. *et al.*, manuscript in preparation). It would be of interest to link these variable CLS phenotypes with modulation in Rsk-2 function. A systematic study of correlations between phenotype and genotype is thus warranted. The lack of information about the exon–intron boundaries of the gene impaired the finding of a larger number of mutations, which at present is 6 out of 76 patients analysed. We are currently revealing the genomic structure of the gene.

Several putative RSK substrates include protein phosphatase-1 (ref. 18) and the transcription factors c-Fos, c-Jun and Nur 77 (ref. 13). CREB also appears to be phosphorylated by Rsk-2 on induction of PC12 cells by nerve growth factor¹⁹. Thus an altered RSK function may influence gene expression by modulating various transcription factors. Indeed, c-Fos is involved in regulating the differentiation and activity of specific bone cell populations during development^{20,21}.

The tissue-specific differences in gene expression suggest distinct physiological roles for the various members of the RSK family^{10,11}.

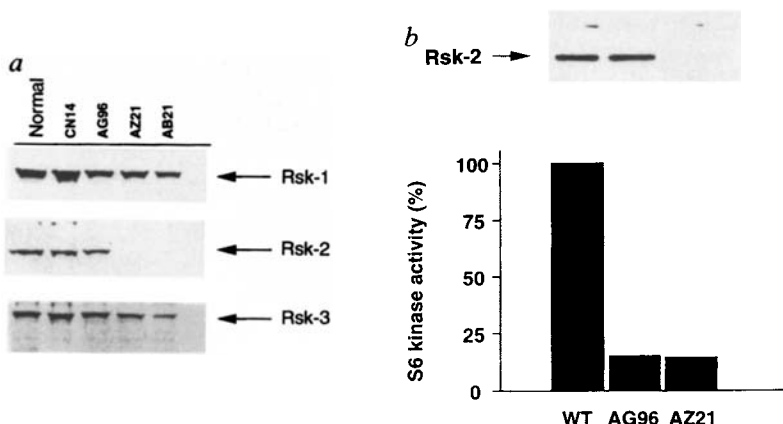


FIG. 4 Rsk-2 proteins from CLS patients are inactive. *a*, Western analysis of RSK protein levels in lymphoblasts from patients. Cell extracts were quantified, and equivalent amounts of total proteins analysed. After blotting, membranes were incubated with specific RSK antibodies. The origin of the lymphoblast cell lines is indicated. *b*, Kinase assay on immunoprecipitated Rsk-2 protein from different lymphoblast cell lines. Top, western analysis of immunoprecipitated Rsk-2 (from 1 mg of total proteins for each lane); bottom, enzymatic activity of the same immunoprecipitated proteins; one unit of Rsk-2 is defined as the amount which catalyses the incorporation of 1 nmol of phosphate into substrate in 1 min. Data are represented as a percentage of Rsk-2 activity immunoprecipitated from the wild-type (WT) cell line. Kinase activity from the AZ21 cell line was reported as a negative control.

Rsk-3 differs with respect to substrate specificity from other RSKs, and may also have distinct upstream activators¹¹. In CLS, Rsk-1 and Rsk-3 are expressed at levels equivalent to normal individuals, indicating that they are not capable of overcoming the Rsk-2 deficiency (Fig. 4a). However, no glycogen metabolism abnormality was found in CLS patients, although Rsk-2 was shown to be responsible for the activation of glycogen synthesis¹⁸. □

Methods

RNA preparation. Total RNA was prepared from lymphoblastoid or fibroblast cell lines using the TRIzol Reagent (GIBCO-BRL)²².

Generation of cDNA probes for Southern blot hybridization. RT-PCR was performed on 1 µg of total RNA from a fibroblast cell line using primers 5'-CTGGCGCAGCTGGCGGA-3' AND 5'-TCAGTTTGAAGCATATGC-3' to amplify a fragment between bases 7 and 420 (as numbered in ref. 5), 5'-CACACT-GAAAGTTCGAGACCG-3' and 5'-CTTTGGGAGTTTTGCAGTA-3' (bases 315–1,102), 5'-GACATTCATTTTCTCAACGA-3' and 5'-GTGCTGAAACAGAAATCCAG-3' (bases 968–1,952), 5'-TGATGATACACAGAGGAAT-3' and 5'-CAAATATCTCACTGAGGTAC-3' (bases 1,866 and 2,245). RT-PCR was performed using the GeneAmp RNA PCR kit (Perkin Elmer). Thermocyclin used the PTC-100 Programmable Thermal Controller (MJ Research) at 94 °C for 5 min followed by 30 cycles of 94 °C for 10 s, 55 °C for 10 s, 72 °C for 10 s, and a final elongation step at 72 °C for 5 min. Random primed P³² probes were hybridized in combination or separately to Southern blots containing genomic DNA from CLS patients, digested using EcoRI or TaqI.

Mutation analysis. Total RNA extracted from fibroblast or lymphoblastoid cell lines established from CLS patients was reverse transcribed using a mixture of the following Rsk-2-derived primers: 5'-TCTATAAATATTCCAGGCT-3' (5' position at 576), 5'-CGTTGAGAAAAATGAATGC-3' (5' position at 987), 5'-TAGTTA-TAGTGAAACGGACAG-3' (5' position at 1,573), 5'-CAAATATCTCACTGAGGTAC-3' (5' position at 2,245). For SSCP analysis, the first-strand cDNA was subjected to PCR amplification using 10 nmol (3 mCi) of ³²P dATP per reaction and the following sets of primers (in separate reactions): 5'-GGACAGCAAAATATGGATGA-3' (5' position at 73) and 5'-AAGCATAATGCACTTGACA-3' (5' position at 399), 5'-GATGTTTCAGAGAGAGAT-3' (5' position at 489) and 5'-TTGGAAAGGGAGTG-TACAG-3' (5' position at 807), 5'-CTAGTGCTAATGCACATCAG-3' (5' position at 1,121) and 5'-AACAGAGTAGGAGCCAACTC-3' (5' position at 1,305) or 5'-TGATAATGTTTGGATGCTGACC-3' (5' position at 1,432), 5'-GCTATGATGCTGC-TTGATG-3' (5' position 1,781) and 5'-ATGCTGAGCACAAGAGCAG-3' (5' position at 2,028). After PCR, amplimers were denatured and loaded onto 10% glycerol, 6% polyacrylamide gels. The gels were run at 5 W for 10–12 h at 4 °C. The PCR-amplified cDNA from patients with altered migration was sequenced directly on an automated DNA sequencer. SSCP analyses of PCR-amplified genomic DNA were performed using the same conditions. The following primer sets gave products of expected size that could be subjected to SSCP analysis: 5'-GACATTCATTTTCTCAACG-3' (5' position at 968) and 5'-CTTTGGGAGTTTTTTCAGTA-3' (5' position at 1,102), 5'-GGTGGTTCATCGAGACTTG-3' (5' position at 1,601) and 5'-CTCTGGTGCAACAAATTTG-3' (5' position at 1,764). In each analysis we tested 80–100 X chromosomes from unrelated normal individuals for nucleotide substitutions.

Western analysis. Lymphoblastoid cells were collected by centrifugation, washed once in cold phosphate-buffered saline (PBS) and lysed in boiling SDS sample buffer. Equivalent amounts of protein from cell lysates were resolved by SDS-PAGE and electroblotted onto Protran nitrocellulose (Schleicher & Schuell). To detect individual isoforms of the RSK proteins, goat polyclonal antibodies specific for Rsk-1 (C-21-G), Rsk-2 (C-19) and Rsk-3 (C-20) (Santa Cruz Biotechnology) were used. Donkey anti-goat-HRP antibodies were used to reveal immune complexes by enhanced chemiluminescence (Amersham).

Immunocomplex kinase assay. Cells were washed once in cold PBS and quickly resuspended in lysis buffer (PBS, 1% Triton X-100, 1% NP40, 0.5% sodium deoxycholate, 0.5 mM DTT, 1 mM Na₃VO₄). After incubation on ice for 10 min, cells were disrupted by repeated aspiration through a 20-gauge needle. The lysates were centrifuged at 13,000 r.p.m. for 15 min at 4 °C and antibodies to Rsk-2 were added to the supernatants. After 2 h of incubation at 4 °C, immune complexes were allowed to react with protein G-Sepharose beads (washed in lysis buffer) for 30 min at 4 °C, then collected by centrifugation and washed three times with lysis buffer and once with kinase buffer (20 mM MOPS, pH 7.2, 25 mM β-glycerol phosphate, 5 mM EGTA, 1 mM Na₃VO₄, 1 mM DTT). Washed immune complexes were divided in two portions, of which one was used for quantification of immunoprecipitated Rsk by western analysis. The other portion was used for kinase assay. Kinase reactions were performed using an S6 kinase assay kit (Upstate Biotechnology).

- Bjorbaek, C. et al. *Diabetes* **44**, 90–97 (1995).
- Trump, D. et al. *Hum. Genet.* **97**, 60–68 (1996).
- Wasserman, D. A., Solomon, N. M. & Rubin, G. M. *Gene* **144**, 309–310 (1994).
- Alcorta, D. A. et al. *Mol. Cell. Biol.* **9**, 3850–3859 (1989).
- Grove, J. R. et al. *Biochemistry* **32**, 7727–7738 (1993).
- Moller, D. E., Xia, C.-H., Tang, T., Zhu, A. & Jakubowski, M. *Am. J. Physiol.* **266**, C351–C359 (1994).
- Zhao, Y. L., Bjorbaek, C., Weremowicz, S., Morton, C. & Moller, D. E. *Mol. Cell. Biol.* **15**, 4353–4363 (1995).
- Jones, S. W., Erikson, E., Blenis, J., Maller, J. & Erikson, R. L. *Proc. Natl Acad. Sci. USA* **85**, 3377–3381 (1988).
- Blenis, J. *Proc. Natl Acad. Sci. USA* **90**, 5889–5892 (1993).
- Hanks, S. K. & Hunter, T. *FASEB J.* **9**, 576–596 (1995).
- Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42–52 (1988).
- Bjorbaek, C., Zhao, Y. & Moller, D. E. *J. Biol. Chem.* **270**, 18848–18852 (1995).
- Fisher, T. L. & Blenis, J. *Mol. Cell. Biol.* **16**, 1212–1219 (1996).
- Dent, P. et al. *Nature* **348**, 302–308 (1990).
- Xing, J., Ginty, D. D. & Greenberg, M. E. *Science* **273**, 959–963 (1996).
- Grigoriadis, A. E. et al. *Science* **266**, 443–448 (1994).
- Hu, E. et al. *EMBO J.* **13**, 3094–3103 (1994).
- Chonczynski, P. *Biotechniques* **15**, 532 (1993).

ACKNOWLEDGEMENTS. We thank A. David, J. Poncin, N. B. Spinner, L. Bason and S. Gilgenkrantz for referring patients; A. Staub, I. Colas, F. Ruffenach, E. Troesch and S. Vicaire for assistance; and M. Chery and Genethon for establishing cell lines. This work was supported by a fellowship from the European Community (D.D.C.), grants from the Commission of the European Communities, the Ministère de la Recherche Française, the Association Française contre les Myopathies, the CNRS and INSERM.

CORRESPONDENCE and requests for materials should be addressed to A.H. (e-mail: andre@titus.u-strasbg.fr).

Requirement of the *paraxis* gene for somite formation and musculoskeletal patterning

Rob Burgess*, Alan Rawls†, Doris Brown†, Allan Bradley‡ & Eric N. Olson†

* Lexicon Genetics Inc., 4000 Research Forest Drive, The Woodlands, Texas 77381, USA

† Department of Molecular Biology and Oncology, Hamon Center for Basic Cancer Research, The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75235, USA

‡ Howard Hughes Medical Institute, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

The segmental organization of the vertebrate embryo is first apparent when somites form in a rostrocaudal progression from the paraxial mesoderm adjacent to the neural tube. Newly formed somites appear as paired epithelial spheres that become patterned to form vertebrae, ribs, skeletal muscle and dermis^{1–3}. Paraxis is a basic helix–loop–helix transcription factor expressed in paraxial mesoderm and somites⁴. Here we show that in mice homozygous for a *paraxis* null mutation, cells from the paraxial mesoderm are unable to form epithelia and so somite formation is disrupted. In the absence of normal somites, the axial skeleton and skeletal muscle form but are improperly patterned. Unexpectedly, however, we found that formation of epithelial somites was not required for segmentation of the embryo or for the establishment of somitic cell lineages. These results demonstrate that *paraxis* regulates somite morphogenesis, and that the function of somites is to pattern the axial skeleton and skeletal muscles.

To determine the function of Paraxis⁴ (also known as bHLH-EC2 (ref. 5) and Meso-1 (ref. 6)), the *paraxis* gene was disrupted using a vector (PTV-1) designed to replace sequences from the initiating methionine codon through the basic helix–loop–helix (bHLH) region and into the first intron with *PGK-neo* (Fig. 1a). Three separate embryonic stem (ES) cell clones heterozygous for the *paraxis* mutation were injected into C57BL/6J blastocysts, and two of the lines yielded chimaeric mice that transmitted the mutant allele through the germ line. These heterozygous mice were phenotypically normal and were bred to homozygosity. The

Received 23 August; accepted 25 October 1996.

- Young, I. D. J. *Med. Genet.* **25**, 344–348 (1988).
- Hanauer, A. et al. *Am. J. Med. Genet.* **30**, 523–530 (1988).
- Biancalana, V. et al. *Genomics* **22**, 617–625 (1994).
- Bird, H., Collins, A. L., Oley, C. & Lindsay, S. *Am. J. Med. Genet.* **59**, 512–516 (1995).

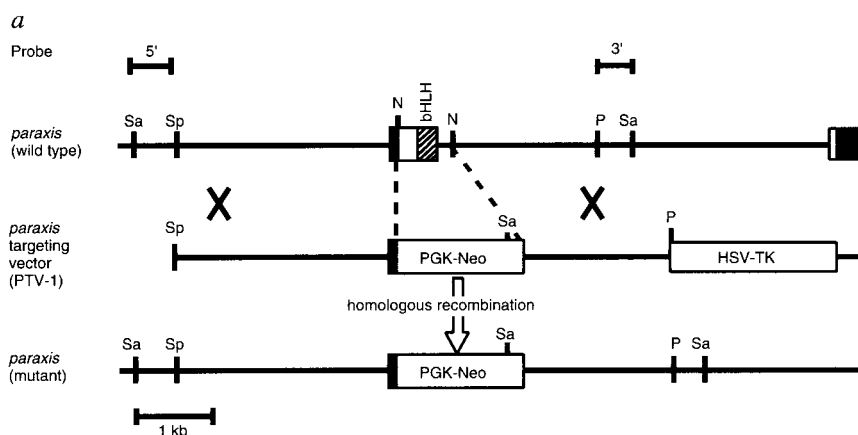
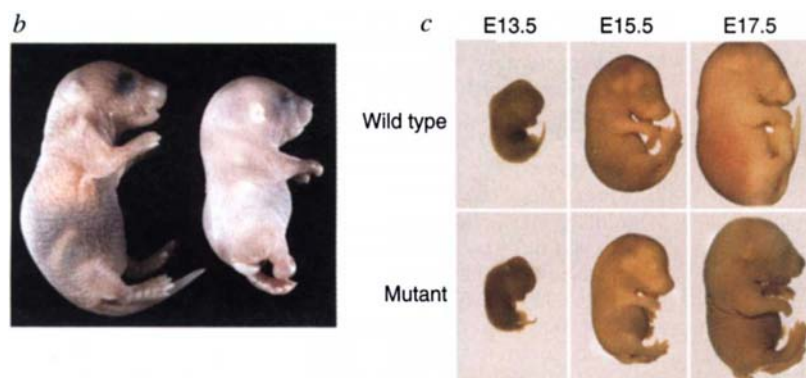


FIG. 1 Targeted disruption of the *paraxis* locus. **a**, Schematic representation of the targeting strategy used to inactivate the *paraxis* gene. The *paraxis* locus contains two exons, separated by a 4.5-kilobase intron. Non-coding regions of the gene are in black, coding regions are in white, and the bHLH region is hatched. A replacement vector, PTV-1, was constructed in which the region from the initiating methionine through the bHLH region into the first intron was deleted and replaced with *PGK-neo*. *HSV-TK* was ligated to the end of the 3' arm of the targeting vector and was used as a negative selection marker. *PGK-neo* introduced a unique *SacI* site that was used to distinguish between the wild-type and targeted alleles by Southern analysis. Restriction sites: *Sa*; *SacI*, *Sp*; *SpeI*, *N*; *NcoI*, *P*; *PstI*. The positions of the 5' and 3' external probes used for Southern analysis are indicated at the top. **b**, Lateral view of wild-type (left) and *paraxis*-null (right) neonates showing severe truncation of the anterior-posterior axis in the mutant. **c**, Wild-type (top) and mutant (bottom) embryos at the indicated days of embryogenesis.



presence of the targeted allele was detected by Southern analysis using probes 5' and 3' of the gene. Genotyped litters from *paraxis* heterozygous intercrosses demonstrated that the targeted allele was transmitted with the predicted mendelian frequency, indicating that the homozygous-null mutation did not result in embryonic lethality. No functional *paraxis* messenger RNA was detected in homozygous mutants by northern blot analysis (data not shown).

At birth, *paraxis*-null neonates showed obvious caudal agenesis with tails that were shortened and curled. Mutant neonates also lacked the normal cervical flexure of the vertebral column, and their hindlimbs were curved towards the midline (Fig. 1b). Mutants also displayed low-set ears, a thickened neck and loose skin. *Paraxis*-null neonates never fed, and died within an hour of birth, possibly because rib defects led to respiratory distress (see below). The posterior truncation of the mutant became apparent at about embryonic day 13 (E13); thereafter, there was little further growth along the posterior axis (Fig. 1c).

To determine whether *paraxis* was required for normal somitogenesis, thin sections through the paraxial mesoderm of wild-type and mutant embryos were stained with haematoxylin and eosin. In sagittal sections of wild-type embryos at E8.5, the rostral somites appeared as well-defined epithelial spheres (Fig. 2a). In contrast, there were no epithelial somites in the paraxial mesoderm of mutant embryos (Fig. 2b); rather, the paraxial mesoderm appeared to be segmented into loose mesenchymal units of approximately the correct size and periodicity of somites. The boundaries between the segmented units of paraxial mesoderm in the mutants were apparent, but were not as sharply demarcated as the intersomitic clefts seen in wild-type embryos. Similarly, transverse sections at E10.5 showed epithelial somites in the proximal tail region of wild-type embryos, but there were no columnar epithelial cells in this region of the mutants (Fig. 2c, d). In more mature somites at later stages, the dermatome, myotome and sclerotome were clearly evident in wild-type embryos, whereas at the same axial level in the mutants, the paraxial mesoderm did not show morphologically identifiable compartments (Fig. 2e, f).

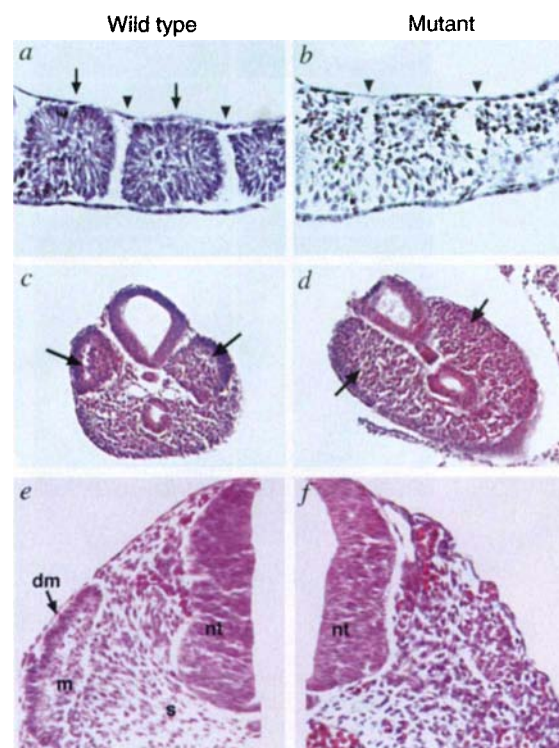


FIG. 2 Histological analysis of somitic regions of wild-type and *paraxis*-null embryos. Embryos sectioned and stained with haematoxylin and eosin. **a**, **b**, Parasagittal sections of E8.5 embryos. The arrows in **a** point to individual epithelial somites and the arrowheads to intersomitic clefts. The arrowheads in **b** point to the clefts between the segmented paraxial mesoderm, which fails to form epithelia. **c**, **d**, Transverse sections through the tail region of E10.5 embryos. The arrows in **c** point to epithelial somites, and those in **d** to paraxial mesoderm without epithelia at a comparable axial level. **e**, **f**, Transverse sections through the caudal tail regions of E11.5 embryos. Wild-type (**a**, **c**, **e**); mutant (**b**, **d**, **f**); dm, dermatomyotome; m, myotome; nt, neural tube; s, sclerotome.

Together, these analyses suggested that *Paraxis* was required for the formation of epithelial somites, but not for segmentation of the paraxial mesoderm.

To determine the fates of cells from the paraxial mesoderm of mutant embryos, we analysed by *in situ* hybridization the expression of several genes specific for different somitic cell lineages. *Pax-1* and *Pax-9* are normally expressed in sclerotomal cells in the ventral region of the somite^{7,8}. In the perinotochordal region of *paraxis* mutant embryos, *Pax-9* (Fig. 3a, b) and *Pax-1* (data not shown) were expressed, but the boundaries of their expression were not as sharply demarcated as in wild-type embryos.

Paraxis shares extensive amino-acid homology with the bHLH protein Scleraxis, which is expressed in chondrogenic precursors of the axial and appendicular skeletons⁹. Scleraxis transcripts were expressed throughout the ventral sclerotomes and rib primordia

of wild-type and mutant embryos (data not shown). Whole-mount *in situ* hybridization of E11.5 embryos also demonstrated that Scleraxis transcripts were expressed in a reiterated pattern along the anteroposterior axis of mutant embryos that appeared to prefigure the ribs (Fig. 3c, d). However, expression of Scleraxis in the dorsal aspect of the paraxial mesoderm was lost in the mutant. The expression of *Pax-1*, *Pax-9* and Scleraxis in mutant embryos demonstrated that the sclerotome can be specified in the absence of epithelial somites.

Pax-7 is normally expressed in the paraxial mesoderm before somite formation¹⁰. As the somite matures, *Pax-7* expression becomes restricted to the myotome, and it later becomes expressed in developing skeletal muscles in the trunk and limbs. *Pax-7* was expressed in the paraxial mesoderm along the anteroposterior axis of mutant embryos, suggesting that myogenic cells were specified without the formation of epithelial somites (Fig. 3e, f). Cells expressing *Pax-7* were located at approximately the correct position relative to the neural tube and notochord, but the boundaries of expression were not sharply defined in the mutants and did not extend as far dorsally and ventrally as in wild-type embryos.

Similar results were obtained with *paraxis* mutant embryos carrying a *myogenin-lacZ* transgene, *Myo1565-lacZ*, which is expressed in the same temporospatial pattern as the endogenous *myogenin* gene¹¹. In mutant embryos at E11.5, the transgene was expressed in the myotome in a metamereric pattern along the anteroposterior axis (Fig. 3g, h). However, the *lacZ*-expressing cells did not assume the dorsoventrally elongated pattern characteristic of the normal myotome. The myogenic bHLH gene *Myf5* and the MADS-box gene *Mef2c*, which encodes a transcription factor important for muscle differentiation¹², were also expressed in the paraxial mesoderm of *paraxis*-null embryos (data not shown), indicating that these genes are regulated independently of *paraxis*.

Pax-3 is expressed in the unsegmented paraxial mesoderm, and during somite differentiation its expression becomes localized to the dermamyotome^{13,14}. *Pax-3* expression was reduced, but was still detectable in the paraxial mesoderm of mutant embryos (Fig. 3i, j). Although the epithelial layer that normally characterizes the dermamyotome was absent in the mutants, the domain of *Pax-3* expression in the paraxial mesoderm corresponded roughly to the position where the dermamyotome should have been. Thus the dermamyotome was molecularly specified despite the inability of these cells to assume their normal epithelial morphology.

To determine whether the abnormal somites in *paraxis*-null embryos resulted in defects in formation or patterning of the axial skeleton, neonates were stained with alizarin red and alcian blue for bone and cartilage, respectively. Despite the absence of morphologically identifiable somites, *paraxis*-null mice had ribs and vertebrae, which, although malformed, showed obvious segmental periodicity (Fig. 4a, b). There were cervical, thoracic and lumbar vertebrae in the mutants, but the vertebral column of *paraxis* mutants lacked the normal cervical flexure, and there was an unsegmented cartilaginous mass that began in the sacral region and extended to the pelvis. In the mutants, it was possible to discern 20–25 vertebral structures, which showed dual ossification centres in the ventromedial region, random fusions in the lateral regions, and malformations (Fig. 4c, d). Posterior to the pelvis, there were no skeletal or cartilaginous elements of the axial skeleton in the homozygous mutants.

Paraxis mutants had ribs, which were underossified, bifurcated and fused, and were detached from the vertebral column (Fig. 4b, f). Several posterior ribs were also missing and the sternums of mutant mice were overossified and lacked intersternal cartilage, which attaches ribs to the sternum. The basioccipital, supraoccipital and exoccipital bones are also derived from the occipital somites, and possibly from cephalic mesoderm¹⁵. The supraoccipital bone in the mutants was hypoplastic (Fig. 4a, b), and the basioccipital and exoccipital bones were fused to each other and to a malformed skeletal element that

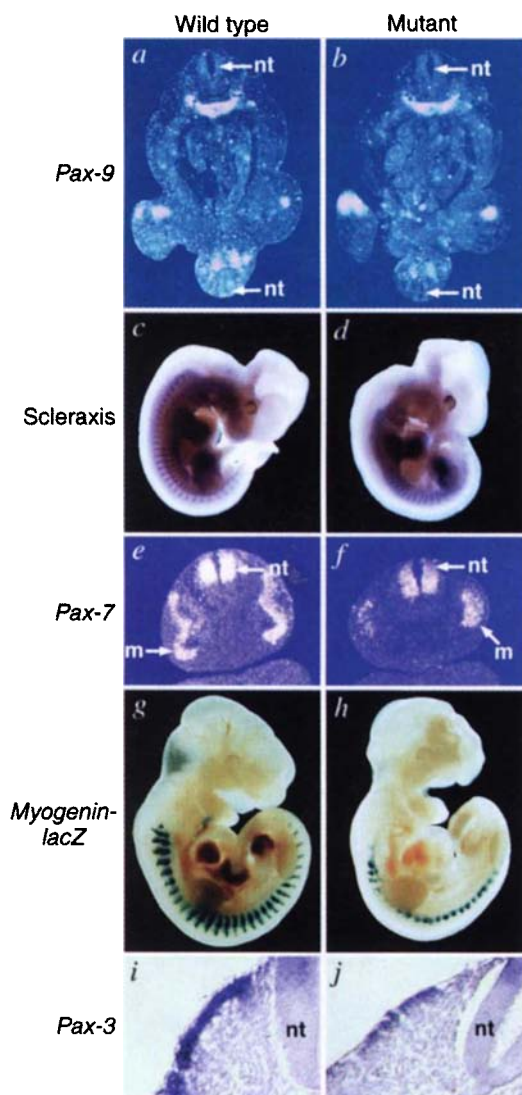


FIG. 3 Expression of somite lineage markers in wild-type and *paraxis*-null embryos. a, b, Detection of *Pax-9* transcripts by *in situ* hybridization to transverse sections through E11.5 embryos at the level of the forelimb. c, d, Detection of Scleraxis transcripts by whole-mount *in situ* hybridization of E11.5 embryos. e, f, Detection of *Pax-7* transcripts by *in situ* hybridization to transverse sections of E11.5 embryos at the level of the proximal tail region. g, h, Detection of *Myo1565-lacZ* expression by whole-mount staining of E11.5 embryos. i, j, Detection of *Pax-3* transcripts in wild-type and mutant embryos at E10.5. Embryos were stained for *Pax-3* transcripts by whole-mount *in situ* hybridization, and embryos were then sectioned transversely. Wild-type, (a, c, e, g, i); mutant (b, d, f, h, j); dm, dermamyotome; m, myotome; nt, neural tube; s, somites.

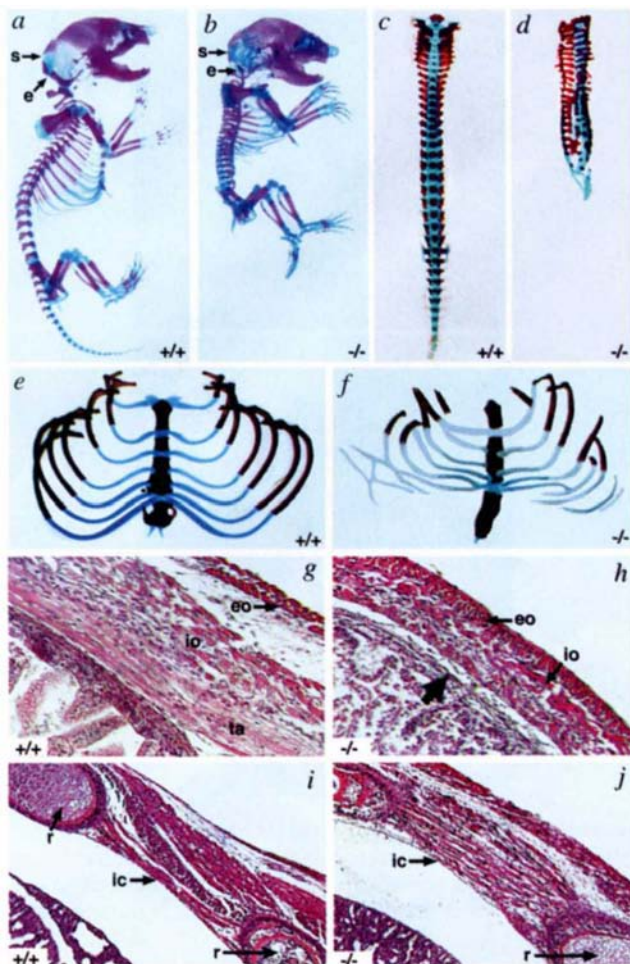


FIG. 4 Skeletal and muscle defects in *paraxis*-null neonates. *a, b*, Wild-type and mutant neonates were stained for bone and cartilage. Genotypes are shown in each panel. Dorsal views of the vertebral columns of wild-type (*c*) and *Paraxis* ($-/-$) (*d*) neonates. *e, f*, Dissected rib cages of wild-type and mutant neonates, respectively. *g, h*, Staining of the abdominal muscle of wild-type (*g*) and mutant (*h*) neonates. The arrow in *h* points to the normal position of the transverse abdominal muscle, which is not detected in the mutant. *i, j*, Staining of transverse sections through ribs and intercostal muscles of wild-type (*i*) and mutant (*j*) neonates. *e*, exoccipital bone; *eo*, external oblique muscles; *ic*, intercostal muscle; *io*, internal oblique muscle; *r*, rib; *s*, supraoccipital bone; *ta*, transverse abdominal muscle.

was probably a remnant of the atlas (Fig. 4*a, b*, and data not shown). Cranial bones derived from cephalic mesoderm and neural crest were unaffected by the *paraxis* mutation. Similarly, the bones of the limbs, pelvis, scapula and clavicle, which are derived from lateral mesoderm, displayed no morphological abnormalities.

Because somites are the source of all skeletal muscle in vertebrate embryos, with the exception of a few muscles of the head, we examined the skeletal muscle of *paraxis* mutant neonates for possible defects. Numerous patterning defects were observed in the axial musculature and the muscles of the limb and body wall, which are derived from the myotome and dermamyotome, respectively². The transverse abdominal muscle of the abdominal wall, for example, was not detected in the mutants (Fig. 4*g, h*). Intercostal muscles were present in the mutants but were hypotrophic and showed aberrant patterning (Fig. 4*i, j*), as did muscles in the limbs (data not shown). These findings demonstrate that formation of

epithelial somites is not required for the development of skeletal muscle, but is essential for appropriate patterning of skeletal muscle fibres. That skeletal muscle was present in the limbs and body wall of the mutants also shows that a correctly formed dermamyotome, from which these muscles are normally derived, is not required to generate the precursors of the lateral musculature.

The phenotype of the *paraxis* mutant mouse demonstrates that *paraxis* is required for the formation of epithelial somites, and exposes a cryptic segmental organization of the embryo that underlies the somites. That the paraxial mesoderm is segmented before somitogenesis has been suggested by stereoscopic scanning electron microscopy, which has revealed spherical clusters of mesenchymal cells called somitomeres that foreshadow the formation of somites^{16,17}, and by the metameric expression pattern of several molecular markers in the paraxial mesoderm^{18–20}. Within the paraxial mesoderm, there are numerous bHLH proteins that regulate morphogenesis, cell determination and differentiation^{9,21–24}. The ability of these different bHLH proteins to interact with common partners and cofactors may allow for precise control of cellular phenotypes as a result of combinatorial interactions with, and competition for, other regulatory proteins. □

Methods

Gene targeting. A targeting vector designated *PTV-1* was generated using pBluescript SK+, in which the neomycin resistance gene replaced the bHLH region of the *paraxis* locus. The herpes simplex virus thymidine kinase gene (*HSV-TK*) was inserted 3' of the targeted locus in the *XhoI* site of the pBluescript polylinker. Electroporation of ES cells and culturing on STO feeders were performed²⁵. After positive-negative selection, surviving clones were isolated and replica plated onto SNL76/7 fibroblast feeder layers in 96-well plates. Southern analysis was performed on these colonies using the probes shown in Fig. 1*a*. ES cell lines heterozygous for a targeted *paraxis* allele were injected into blastocysts isolated from C57BL/6J mice to generate chimaeras which transmitted the mutant allele through the germ line.

Histology and in situ hybridization. Sectioning of embryos and staining with haematoxylin and eosin were performed²⁶. Fixation, embedding, sectioning and *in situ* hybridization of mouse embryos was performed²⁶. For detection of *Myo1565-lacZ* expression, transgenic mice bearing the transgene *Myo1565-lacZ*¹¹ were bred with *paraxis*-heterozygous mice, and offspring heterozygous for the targeted *paraxis* allele and carrying *Myo1565-lacZ* were bred to *paraxis*-heterozygous mice. Transgene expression in the resulting embryos was detected¹¹. Staining was with alizarin red and alcian blue to reveal bone and cartilage, respectively, and staining of embryo sections was with haematoxylin and eosin²⁶.

Received 18 September; accepted 21 October 1996.

- Christ, B. & Ordahl, C. P. *Anat. Embryol.* **191**, 381–396 (1995).
- Ordahl, C. P. & LeDouarin, N. M. *Development* **114**, 339–353 (1992).
- Keynes, R. J. & Stern, C. D. *Development* **103**, 413–429 (1988).
- Burgess, R., Cserjesi, P., Ligon, K. L. & Olson, E. N. *Dev. Biol.* **168**, 296–306 (1995).
- Quertermous, E. E., Hidai, H., Blannar, M. A. & Quertermous, T. *Proc. Natl Acad. Sci. USA* **91**, 7066–7070 (1994).
- Blannar, M. A. et al. *Proc. Natl Acad. Sci. USA* **92**, 5870–5874 (1995).
- Deutsch, U., Dressler, G. R. & Gruss, P. *Cell* **53**, 617–625 (1988).
- Neubuser, A., Koseki, H. & Balling, R. *Dev. Biol.* **170**, 701–716 (1995).
- Cserjesi, P. et al. *Development* **121**, 1099–1110 (1995).
- Jostes, B., Walther, C. & Gruss, P. *Mech. Dev.* **33**, 27–38 (1990).
- Cheng, T.-C., Wallace, M., Merlie, J. P. & Olson, E. N. *Science* **261**, 215–218 (1993).
- Martin, J. F., Schwarz, J. J. & Olson, E. N. *Proc. Natl Acad. Sci. USA* **90**, 5282–5286 (1993).
- Goulding, M. et al. *Genomics* **17**, 355–363 (1993).
- Williams, B. A. & Ordahl, C. P. *Development* **120**, 785–796 (1994).
- Couly, G. F., Coitey, M. & LeDouarin, N. M. *Development* **117**, 409–429 (1993).
- Jacobson, A. G. *Development* **104**, 209–220 (1988).
- Tam, P. P. L. & Meir, S. *Am. J. Anat.* **164**, 209 (1982).

- Nieto, M. A., Gilardi-Hebenstreit, P., Chamay, P. A. & Wilkinson, D. G. *Development* **116**, 1137–1150 (1992).
- Yamaguchi, T. P., Conlon, R. A. & Rossant, J. *Dev. Biol.* **152**, 75–88 (1992).
- Muller, M., Weizsacker, E. & Campos-Ortega, J. A. *Development* **122**, 2071–2078 (1996).
- Olson, E. N. & Klein, W. H. *Genes Dev.* **8**, 1–8 (1994).
- Fan, C.-M. & Tessier-Lavigne, M. *Cell* **79**, 1175–1186 (1994).
- Li, L., Cserjesi, P. & Olson, E. N. *Dev. Biol.* **172**, 280–292 (1995).
- Wolf, C. et al. *Dev. Biol.* **143**, 363–373 (1991).
- McMahon, A. & Bradley, A. *Cell* **62**, 1073–1085 (1990).
- Martin, J. F., Bradley, A. & Olson, E. N. *Genes Dev.* **9**, 1237–1249 (1995).

ACKNOWLEDGEMENTS. We thank P. Tam, W. Wright and M. Perry for comments; A. Tizenor for assistance with graphics; T. Davis for editorial assistance; R. Balling for Pax-9; P. Gruss for Pax-7; and C. Ordahl for Pax-3. This work was supported by grants from the NIH, The Muscular Dystrophy Association, and the Human Science Frontiers Program (E.N.O.). Part of this work was done while R.B. and D.B. were graduate students at the University of Texas M. D. Anderson Cancer Center. A.R. was supported by a postdoctoral fellowship from the Muscular Dystrophy Association. A.B. is an associate investigator with the Howard Hughes Medical Institute.

CORRESPONDENCE and requests for materials should be addressed to E.N.O. (e-mail: eolson@hamon.swmed.edu).

Requirement for Brn-3.0 in differentiation and survival of sensory and motor neurons

Robert J. McEvelly*, Linda Erkman*, Lin Luo†, Paul E. Sawchenko‡, Allen F. Ryan† & Michael G. Rosenfeld*

* Howard Hughes Medical Institute, University of California, San Diego, Department and School of Medicine, and † Departments of Surgery/Otolaryngology and Neurosciences, and VA Medical Center, 9500 Gilman Drive, La Jolla, California 92093, USA

‡ Laboratory of Neuronal Structure and Function, The Salk Institute for Biological Sciences, P.O. Box 85800, San Diego, California 92186, USA

SPECIFIC families of transcription factors mediate events in the sequential maturation of distinct neuronal phenotypes. Members of one such family, the class IV POU domain transcription factor Brn-3.0, and two highly related factors Brn-3.1 and Brn-3.2, are differentially expressed in the developing and mature mammalian nervous system^{1–11}. The expression pattern of Brn-3.0 suggested that it has an important role in the development of sensory ganglia, as well as red nucleus, inferior olive, and nucleus ambiguus. Analysis of mice null for the *Brn-3.0* locus shows that Brn-3.0 is required for the survival of subpopulations of proprioceptive, mechanoreceptive and nociceptive sensory neurons, where deletion of the gene affects neurotrophin and neurotrophin-receptor gene expression. Deletion of *Brn-3.0* also alters either differentiation, migration or survival of specific central neuronal populations.

Expression of the gene encoding Brn-3.0 co-distributes with that of the low-affinity nerve growth factor receptor (p75) in adjacent sections of presumptive migrating neural crest of the mouse at embryonic day 9.5 (E9.5) (Fig. 1), and persists throughout development of the sensory nervous system. In the central nervous system it is restricted to discrete regions of the midbrain, hindbrain and spinal cord^{6,9–11}. To assess the role of Brn-3.0, the entire transactivating and critical DNA binding regions were deleted by targeted homologous recombination in embryonic stem cells¹². This generated apparently normal mice heterozygous for the null mutation and homozygous null mice at slightly less than mendelian ratios (18/100), with null mice exhibiting the expected gene deletion and absence of Brn-3.0 protein (Fig. 2a–c; see Supplementary Information). Although homozygous null mice *Brn-3.0*^{−/−} seemed to be grossly normal, they exhibited

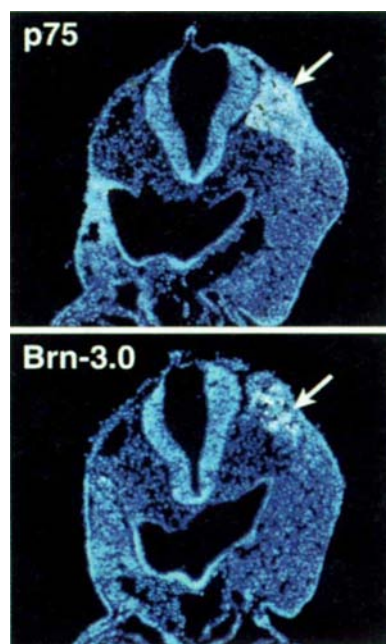


FIG. 1 Expression of low-affinity nerve growth factor receptor (p75) and *Brn-3.0* mRNA in adjacent transverse sections of a 13-somite-stage mouse embryo shown by *in situ* hybridization²⁹.

severe behavioural and physiological problems, including ineffective swallowing and an inability to right themselves, intermittently displaying abnormal movement and posture including rigid outstretched limbs and head; they survived less than 24 h postpartum.

These severe behavioural problems and expression of Brn-3.0 throughout the development of central nervous system structures related to movement and feeding prompted us to examine brainstem nuclei. Histological analysis of the *Brn-3.0*^{−/−} hindbrain at postpartum day 0.5 (P0.5) revealed an absence of red nucleus and compact formation nuclei, and a loss of multiple subunits of the inferior olive (Fig. 3). *Brn-3.0*^{−/−} neurons could be visualized because the gene replacement vector generated a transcript under control of the Brn-3.0 promoter containing the distal 3'-untranslated region of Brn-3.0. A probe specific for a portion of this transcript (referred to as 3M) also hybridized to wild-type *Brn-3.0* messenger RNA. In the red nucleus, the 3M probe revealed the presence of the entire Brn-3.0 neuronal population at E16.5,

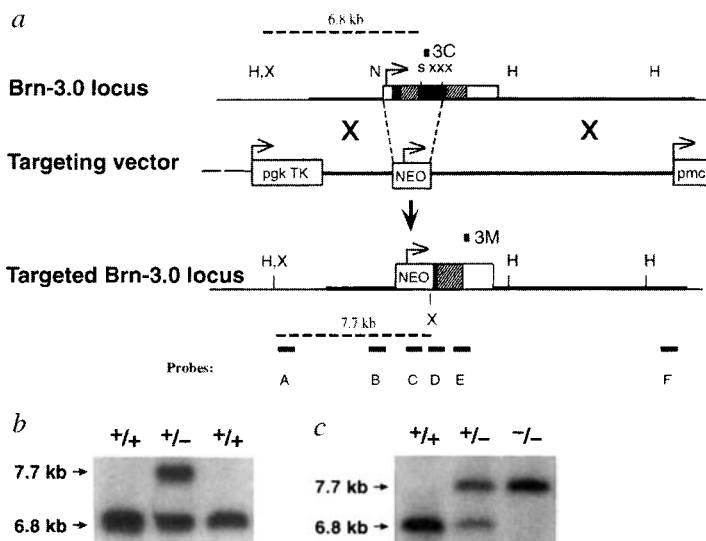


FIG. 2 Generation and analysis of *Brn-3.0*^{−/−} mice. **a**, *Brn-3.0* gene, disrupted allele and diagnostic probes. Filled, open and hatched boxes represent coding exons, non-coding exons, and introns, respectively. H, *HindIII*; N, *NotI*; S, *Sall*; X, *XhoI*. Probes: 3C; nucleotides 833–1,058 (amino acids 189–264) of Brn-3.0, 3M; nucleotides 1,533–2,037, (A); external, (B–F); internal. **b**, Southern blot analysis of transfected cell lines with probe A following *HindIII/Sall* digestion yields 7.7-kb mutant fragment in one cell line. **c**, Southern blot analysis of a litter generated by the intercross of *Brn-3.0*^{+/−} mice.

with progressive loss at E18.5, such that by P0.5 only a few scattered 3M-positive cells could be detected (data not shown). These results suggest that red nucleus neurons are born, migrate to their appropriate positions in early development, but subsequently fail to survive between E16.5 and P0.5. A similar loss of red nucleus neurons has been observed with axotomy of the rubrospinal tract, and can be rescued with the exogenous application of the neurotrophic factor BDNF¹³, suggesting that *Brn-3.0* might be required for either the response to, or synthesis of, BDNF.

In contrast, the ontogenic analysis of markers for the inferior olive and nucleus ambiguus, formed by successive cellular migrations, suggests that these structures lack specific subunits owing to deficiencies in specific phases of developmental migrations. Examination of the inferior olive, generated by migrations of cells from the inferior olivary migratory stream between E15 and P0, revealed the expected arrangement of subunits at E16.5, but a distortion of several of the component units that segregate into final positions between E18 and P0 (ref. 14) (Fig. 3b and data not shown). Most obvious was the loss of portions of the anterolateral inner and outer lamellae of the principal olive, a substrate extending through the rostral half of the nucleus (Fig. 3b), but some dorsolateral subunits of the caudal half of the nucleus were also altered or absent¹⁴ (data not shown). In the nucleus ambiguus the motor neurons of the compact formation innervate the oesophagus, and are required for the peristaltic phase of the swallowing reflex¹⁵. The absence of the compact formation in histological analysis of mutant brains at P0.5 (Fig. 3b) was further documented in an *in situ* hybridization analysis of the *Brn-3.0*^{-/-} hindbrain at E16.5, E18.5 and P0.5 with the 3M probe, and with probes for α -CGRP and choline acetyl transferase (ChAT), two markers of terminal differentiation that normally codistribute with *Brn-3.0* in the compact formation (Fig. 3c, and data not shown). However, neurons that would have expressed *Brn-3.0* (3M-positive) were identified caudal to the compact formation and dorsal to the semicompact and loose formations of the ambigular complex, in the mutant, but not wild-type, hindbrains (Fig. 3d). These ectopic cells failed to express α -CGRP and ChAT (Fig. 3d, and data not shown). Our data suggest that, in the absence of *Brn-3.0* gene expression, the compact formation neurons, which originate from the midline neuroepithelium in the fourth rhombomere at E10 to E11, mis-migrate to assume a position dorsal to the semicompact formation, rather than to the usual ventrolateral distribution in the rostral nucleus ambiguus in the sixth rhombomere at E14 to 15 (ref. 16) (Fig. 3e).

Ontogenic and postnatal analyses of cranial and dorsal root ganglia in *Brn-3.0*^{-/-} mice revealed substantially reduced expression of biochemical markers characteristic of sensory neuron development. By P0.5 there was also a pronounced loss of cellularity in the trigeminal, acoustic and cervical dorsal root ganglia. Light microscopy of semithin plastic sections at P0.5 demonstrated a distortion in the migration of acoustic ganglion sensory neurons which normally occurs between E14 and E17 during the outgrowth of the cochlea¹⁷; neurons were scattered throughout the nerve in the modiolus (centre) of the cochlea, and there was a loss of approximately 50% of the normal complement of neurons (Fig. 4a, and data not shown). *In situ* hybridization

analysis of these cells with the probe 3M revealed labelling over the entire vestibular and cochlear ganglia in the *Brn-3.0*^{-/-} mice, although *Brn-3.0* (3C) and *Brn-3.2* transcripts were not detected (Fig. 4a, and data not shown). Electron micrographs revealed the presence of terminals of cochlear and vestibular neurons on inner hair cells of the cochlear duct and on vestibular hair cells in the *Brn-3.0*^{-/-} animals (Fig. 4b), indicating that the axons of these cells can reach their peripheral targets.

The trigeminal ganglia of these *Brn-3.0*^{-/-} mice were of normal size from E12.5 to E16.5, although by P0.5 they exhibited an almost complete loss of the dorsal division of the anteromedial lobe of the ganglion¹⁶ (Fig. 4c), with remaining neurons expressing the 3M probe (see Supplementary Information). At P0.5, *in situ*

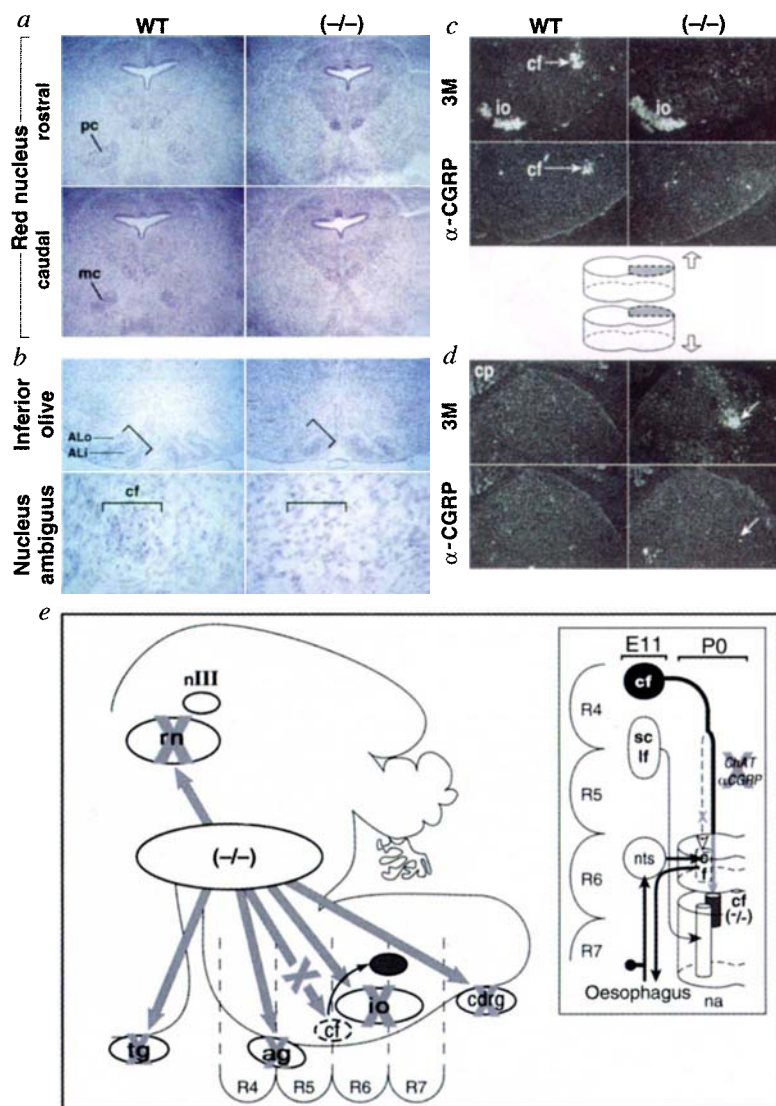


FIG. 3 Central nervous system defects in *Brn-3.0*^{-/-} mice. **a**, Coronal sections of paraffin-embedded wild-type (WT) and mutant P0.5 brains ($-/-$); parvocellular (pc) and magnocellular (mc) nuclei of the red nucleus are indicated. **b**, Top, brackets demonstrate the absence of the anterolateral inner and outer lamellae (ALo and ALI) of the principal olivary nucleus; bottom, a high-magnification view of the same region demonstrating the absence of the nucleus ambiguus compact formation (cf). **c**, *In situ* hybridization with 3M and α -CGRP cRNA probes in coronal sections through the ventrolateral medulla at the level of the cf; io, inferior olive. **d**, *In situ* hybridization through the dorsomedial medulla at the level of the nucleus ambiguus semicompact and loose formations; cp, choroid plexus. **e**, Defects in the *Brn-3.0*^{-/-} nervous system, and reflex arc connecting oesophageal sensory afferents, the nucleus of the solitary tract (nts) and cf motor neurons; ag, acoustic ganglion; cdrg, cervical dorsal root ganglia; ChAT, choline acetyl transferase; lf, loose formation; nIII, motor nucleus III; sc, semicompact formation; tg, trigeminal ganglion; R4–R7, rhombomeres 4–7.

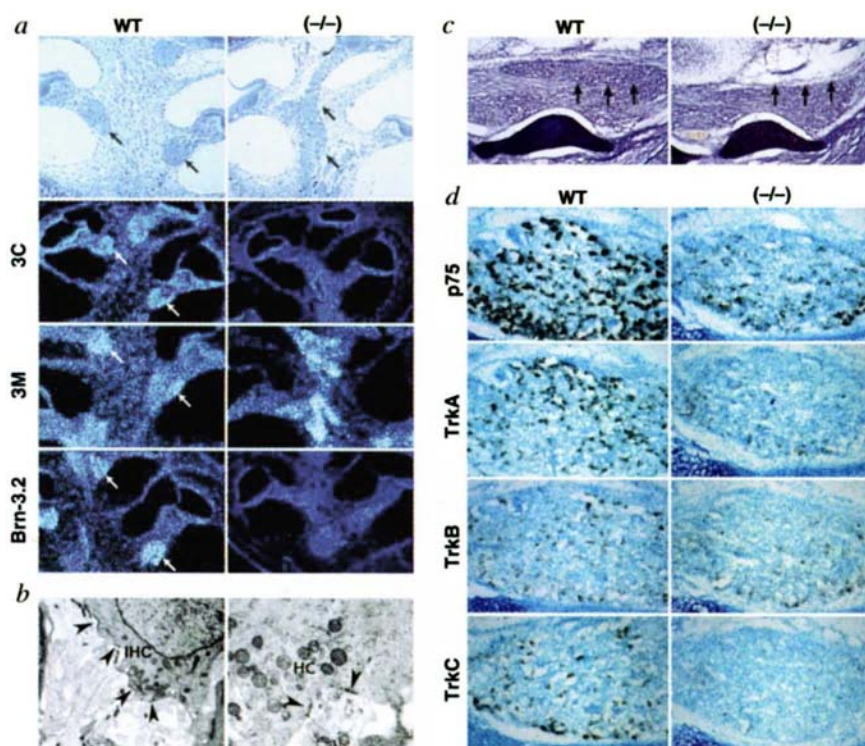


FIG. 4 Analysis of cranial sensory ganglia. *a*, Semithin plastic sections³⁰ and *in situ* hybridization of wild-type (WT) and mutant ($-/-$) acoustic ganglia of newborn cochlea; arrows indicate acoustic ganglion. *b*, Electron microscopy of hair cells in cochlea and saccular macula of a P0.5 *Brn-3.0*^{-/-} mouse. Left, cochlear inner hair cell (IHC); right, vestibular hair cell (HC). Arrows indicate apparent synapses between hair cells and sensory neurons. *c*, Arrows indicate near-complete loss of dorsal division of the anteromedial lobe in Nissl-stained parasagittal sections of P0.5 mutant trigeminal ganglia. *d*, *In situ* hybridization of adjacent coronal sections of P0.5 trigeminal ganglia in wild-type and mutant mice with p75, TrkA, TrkB and TrkC probes.

hybridization analysis of adjacent sections in wild-type and mutant ganglia revealed a substantial reduction in both the number of p75-, TrkA-, TrkB-, and TrkC-expressing neurons per unit area and the levels of expression of these markers (Fig. 4*d*), and expression of BDNF was undetectable (see Supplementary Information). *Brn-3.1* and *Brn-3.2* transcripts were not detected in the *Brn-3.0*^{-/-} ganglion, despite widespread labelling with two other transcription factors, *Isl-1* (ref. 18) and *NZF-1* (ref. 19) (see Supplementary Information).

In dorsal root ganglia of *Brn-3.0*^{-/-} mice, similar results were observed at the cervical level at P0.5, including a reduction in TrkA and TrkB, an absence of TrkC expression in the rostral cervical ganglia, and an absence of BDNF, despite vigorous expression of the markers 3M and *Isl-1* in adjacent sections (Fig. 5*a, b, d* and data not shown; see Supplementary Information). DiI fluorescent tracing²⁰ of the central projections of dorsal root ganglia (c1–2) revealed a lack of proprioceptive projections to the ventral horn and vastly reduced numbers of projections to laminae I–IV (ref. 21), whereas a normal pattern of spinal innervation was observed for the c5–8, thoracic and lumbar levels (Fig. 5*c*, and data not shown).

An ontogenic analysis demonstrated that deletion of the gene encoding *Brn-3.0* altered early patterns of gene expression in both NGF-dependent and NGF-independent populations of the trigeminal and dorsal root ganglia. Reductions in p75, TrkB and TrkC expression levels were detected as early as E12.5 in trigeminal ganglia, apparent decreases in the number of TrkC-positive cells were detected at E16.5, and decreases in the numbers of TrkA- and TrkB-positive cells were apparent on P0.5 (see Supplementary Information). Differences in cell number and expression levels were initially obvious in the cervical regions, and observed only at later times in the thoracic and lumbar

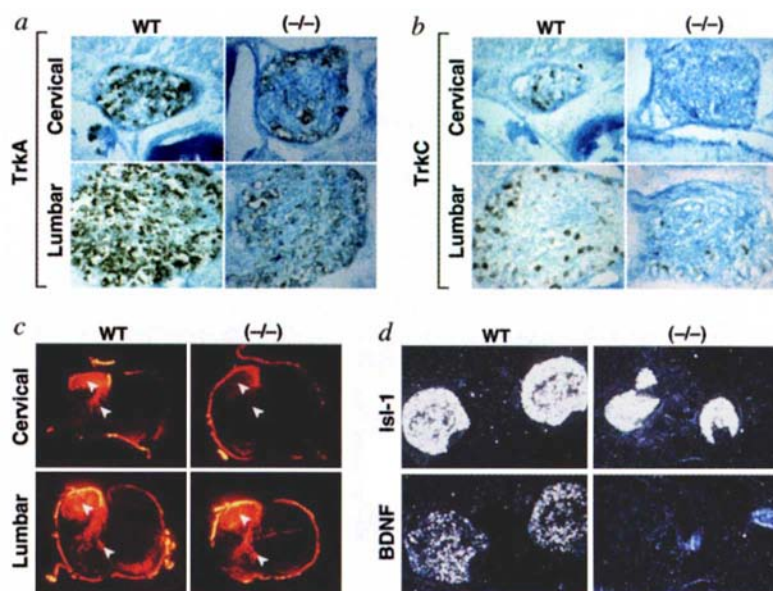


FIG. 5 Defects in sensory ganglia development. *a, b, d*, *In situ* hybridization of adjacent sections of P0.5 cervical (*a*), lumbar (*b*) and thoracic (*d*) dorsal root ganglia of wild-type (WT) and mutant ($-/-$) mice with TrkA and TrkC or BDNF and *Isl-1* probes. *c*, Arrows indicate sensory neuron projections from the dorsal root to dorsal laminae I–IV and ventral horn in the cervical and lumbar spinal cord of P0.5 WT and mutant ($-/-$) mice following DiI tracing.

levels, demonstrating a rostral to caudal gradient of *Brn-3.0*-dependent events in the sensory nervous system (see Supplementary Information).

Together, the defects observed in *Brn-3.0*^{-/-} mice would account for the motor and ingestive behavioural phenotypes, which are similar to those of mice deficient in TrkB and BDNF^{22–25}. The observed deficiencies in these *Brn-3.0*^{-/-} mice suggest that an aspect of neurotrophin-mediated signalling might

be responsible for the failure of sensory²⁶ and red-nucleus neurons to survive. Although expression of Brn-3.0 commences early in neural development, functional consequences of its deletion become apparent only relatively late in neuronal maturation. This is analogous to the deficiencies caused by the deletion of genes encoding Brn-3.1 and Brn-3.2 (refs 27, 28), and might be due to a requirement for Brn-3.0 in programmed patterns of gene expression, the execution of cellular migrations, and neuronal survival. Thus our data on mice with deleted genes for Brn-3.0, Brn-3.1 and Brn-3.2 provide a link between class IV POU domain transcription factors and the development and survival of distinct mammalian neurons and sensory cells.

Note added in proof: Results of a similar deletion experiment have just been reported by Xiang *et al.* in *Proc. Natl Acad. Sci. USA* **93**, 11950–11955 (1996). □

Methods

Targeted disruption of the genomic locus. A NotI–XhoI fragment coding for amino acids 1–370 of the protein was deleted from the mouse *Brn-3.0* locus, isolated from a library of 129/sv mouse DNA (Stratagene), and replaced with a *pgk-neo* cassette. Expression of the neomycin gene was driven by the mouse phosphoglycerate kinase (*pgk*) promoter. The targeting vector consisted of 3.3 kilobases of 5' homologous sequences and 10 kb of 3' homologous sequences, and resulted in the positioning of the negative selection thymidine kinase (*TK*) gene at both ends of the homologous *Brn-3.0*-flanking DNA regions. After correct targeting, amino acids 1–370 were replaced with the *pgk-neomycin* cassette. Embryonic stem (ES) cells were transfected and selected^{12,29}. Probe A (Fig. 2) was used to detect homologous recombination in transfected ES cell lines, and probes B–F were used with *Sall*, *NotI*, *HindIII* and *EcoRI* restriction digests to confirm results obtained with probe A (data not shown).

Immunohistochemical, histological and *in situ* hybridization studies. A polyclonal antibody, raised against a fusion peptide encompassing amino acids 189–264 of Brn-3.0 protein, was used for immunohistochemistry, after preadsorption to adult mouse cortex²⁷. For paraffin sections, P0.5 mice were perfused with Bouin's fixative, dissected brains were postfixed and processed²⁹, and 10- μ m serial sections were stained with cresyl violet. For semithin sections, plastic-embedded tissues were prepared from P0.5 mice and semithin sections prepared and stained with toluidine blue³⁰. Tissues were prepared and subjected to electron microscopy³⁰. For *in situ* hybridization, tissues were processed and hybridized according to ref. 29, except for studies of ear²⁷. cRNA probes were derived from the third exon of mouse *p75* cDNA (gift from J. Huber), from the fifth and sixth exons of mouse α -CGRP, and from cDNAs encoding the extracellular domains of mouse TrkA, TrkB and rat TrkC (gift from M. E. Palko), the 3'-untranslated region of mouse Brn-3.2, the N termini of rat NZF-1 and mouse Isl-1 and Brn-3.1, and the coding region of rat BDNF (gift from F. Gage). The 3C and 3M probes were derived from mouse *Brn-3.0* cDNA, as described in Fig. 2.

Dil tracing of dorsal root ganglia projections into spinal laminae. Dil tracing was performed as described in ref. 20, with slight modifications. *Brn-3.0*^{+/+} and *Brn-3.0*^{-/-} P0.5 mice were perfused with 4% paraformaldehyde and postfixed overnight at room temperature. Spinal cords with attached ganglia were dissected, the attached ventral roots cut, and individual ganglia injected with 0.25% Dil (1,1'-diocetadecyl-3,3',3'-tetramethylindocarbocyanine perchlorate) in dimethylformamide. The labelling was allowed to proceed for 2 weeks in 4% paraformaldehyde at 37°C, and 20- μ m cryostat sections were viewed with a rhodamine filter using a Zeiss Axioscop microscope.

Received 12 August; accepted 28 October 1996.

1. Finney, M., Ruvkun, G. & Horvitz, H. R. *Cell* **55**, 757–769 (1988).
2. He, X. *et al. Nature* **340**, 35–41 (1989).
3. Collum, R. G. *et al. Nucleic Acids Res.* **20**, 4919–4925 (1992).
4. Lillycrop, K. A. *et al. Nucleic Acids Res.* **20**, 5093–5096 (1992).
5. Bhargava, A. K., Li, Z. & Weissman, S. M. *Proc. Natl Acad. Sci. USA* **90**, 10260–10264 (1993).
6. Gherro, M. R. *et al. Proc. Natl Acad. Sci. USA* **90**, 10841–10848 (1993).
7. Ninkina, N. N., Stevens, G. E., Wood, J. N. & Richardson, W. D. *Nucleic Acids Res.* **21**, 3175–3182 (1993).
8. Xiang, M. *et al. Neuron* **11**, 689–701 (1993).
9. Turner, E. E., Jenne, K. J. & Rosenfeld, M. G. *Neuron* **12**, 205–218 (1994).
10. Xiang, M. *et al. J. Neurosci.* **15**, 4762–4785 (1995).
11. Fedtsova, N. G. & Turner, E. E. *Mech. Dev.* **53**, 291–304 (1995).
12. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. *Nature* **336**, 348–352 (1988).
13. Diener, P. S. & Bregman, B. S. *NeuroReport* **5**, 1913–1917 (1994).
14. Altman, J. & Bayer, S. A. *J. Comp. Neurol.* **257**, 490–512 (1987).
15. Cunningham, E. T. Jr & Sawchenko, P. E. *J. Neurosci.* **9**, 1668–1682 (1989).
16. Altman, J. & Bayer, S. A. *Adv. Anat. Embryol. Cell Biol.* **74**, 1–90 (1982).
17. Sher, A. E. *Acta Otolaryngol. (Stockh.)* (suppl.) **285**, 1–77 (1971).
18. Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. & Edlund, T. *Nature* **344**, 879–882 (1990).
19. Jiang, Y. *et al. J. Biol. Chem.* **271**, 10723–10730 (1996).
20. Ernfor, P., Lee, K. F., Kucera, J. & Jaenisch, R. *Cell* **77**, 503–512 (1994).
21. Ruit, K. G., Elliott, J. L., Osborne, P. A., Yan, Q. & Snider, W. D. *Neuron* **8**, 573–587 (1992).
22. Klein, R. *et al. Cell* **75**, 113–122 (1993).
23. Ernfor, P., Lee, K. & Jaenisch, R. *Nature* **368**, 147–150 (1994).

24. Jones, K. R., Farifias, I., Backus, C. & Reichardt, L. F. *Cell* **76**, 989–999 (1994).
25. Conover, J. C. *et al. Nature* **375**, 235–238 (1995).
26. Buchman, V. L. & Davies, A. M. *Development* **118**, 989–1001 (1993).
27. Erkman, L. *et al. Nature* **381**, 603–606 (1996).
28. Gan, L. *et al. Proc. Natl Acad. Sci. USA* **93**, 3920–3925 (1996).
29. Schonemann, M. D. *et al. Genes Dev.* **9**, 3122–3135 (1995).
30. Schwartz, I. R. & Ryan, A. F. *J. Electron Microsc. Tech.* **15**, 225–244 (1990).

SUPPLEMENTARY INFORMATION is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

ACKNOWLEDGEMENTS. We thank S. O'Connell for help with studies of neural crest; A. Ryan for the Brn-3.0 antibody; M. Ayers for help with developing the founder mice; T. Bennet for electron microscopy; K. Kalla for microinjection of blastocysts; E. Mears for genotype analysis; E. Li for J1 ES cells; and P. Myer for preparing illustrations. L.E. was supported by the Swiss National Science Foundation. P.E.S. is an investigator of the Foundation for Medical Research. M.G.R. is an investigator with the Howard Hughes Medical Institute. This research was supported by grants from NIH/NINDS (P.E.S. and M.G.R.) and by NIH/NIDCD and VA Research Service (A.F.R.).

CORRESPONDENCE and requests for materials should be addressed to M.G.R. (e-mail: mrosenfeld@ucsd.edu).

CD8 enhances formation of stable T-cell receptor/MHC class I molecule complexes

K. Christopher Garcia*, Christopher A. Scott*, Anders Brunmark, Francis R. Carbone†, Per A. Peterson, Ian A. Wilson* & Luc Teyton

The R. W. Johnson Pharmaceutical Research Institute, La Jolla, 3535 General Atomic Court, San Diego, California 92121, USA

* The Scripps Research Institute and the Skaggs Institute for Chemical Biology, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

† Monash University, Alfred Hospital Prahran, Victoria 3181, Australia

T-CELL antigen receptors (TCR) generally interact with moderate affinity with the complex formed by major histocompatibility complex (MHC) molecules and foreign peptides^{1–7}. MHC/TCR recognition is followed by the generation of a signal to the T cell through a monomorphic multicomponent system that includes the CD3 complex and accessory molecules such as CD4 and CD8. The interaction between the extracellular domains of MHC and TCR molecules^{1–7}, and the interaction of MHC and CD4/CD8 molecules^{8–10}, have been considered to occur independently of one another. We report here that the affinity of CD8 dimers for MHC class I molecules is independent of haplotype and peptide content, and that the affinity of the TCR for its specific ligand is enhanced through a reduced 'off' rate in the presence of either CD8 $\alpha\alpha$ homo- or CD8 $\alpha\beta$ heterodimers. Moreover, CD8 seems to help recognition of the specific MHC-peptide complex either by guiding an energetically favourable docking of TCR onto MHC, or by inducing conformational changes in the MHC complex that can augment the TCR/MHC-peptide interaction. CD8 should therefore be considered as an active participant in the T-cell recognition complex, rather than simply as an accessory molecule.

Soluble, secreted MHC class I molecules were expressed in *Drosophila melanogaster* cells without addition of specific peptide adducts¹¹. To produce soluble heterodimeric $\alpha\beta$ TCR, we deleted the sequence coding for the transmembrane segment of each chain from the complementary DNA, but preserved the sequences encoding the cytoplasmic tail. We confirmed the 1:1 $\alpha:\beta$ chain stoichiometry of the purified material by N-terminal protein sequencing. We then checked the native conformation of the purified TCRs using a panel of anti-TCR antibodies in a surface plasmon resonance (SPR) assay. We engineered and expressed soluble forms of CD8 $\alpha\alpha$ and CD8 $\alpha\beta$ dimers, and checked the 1:1 stoichiometry of the α - and β -chains in the CD8 $\alpha\beta$ heterodimer using similar strategies. We analysed size homogeneity at high concentration by gel filtration. TCR, MHC and CD8 molecules, as

expected, had relative molecular masses (M_r s) of 56K, 47K and 46K, respectively, and did not multimerize or aggregate, allowing accurate affinity measurements by SPR, which is based on mass determination¹².

Purified CD8 $\alpha\beta$ and CD8 $\alpha\beta$ dimers were used to measure

MHC/CD8 interactions by SPR (Fig. 1). CD8 injections over the MHC class I K^b-OVA surface resulted in measurable and specific binding. Reliable kinetic rate constants can be obtained for higher CD8 concentrations, in spite of a baseline shift that occurs as an occasional artefact when studying low-affinity inter-

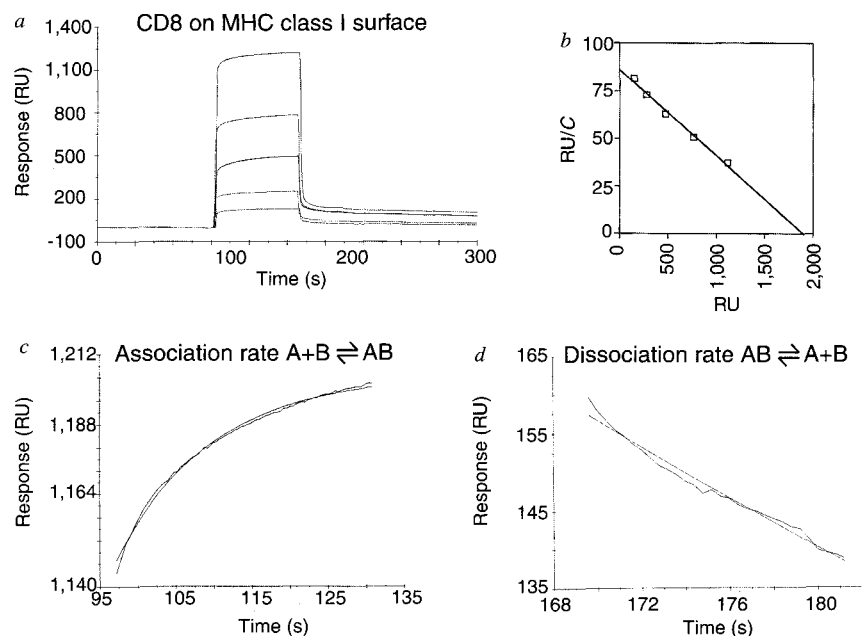


FIG. 1 CD8 $\alpha\beta$ dimers interact with immobilized MHC class I molecules. *a*, BIAcore sensorgrams of CD8 over a H-2 K^b-OVA surface. *b*, Resonance unit (RU)/concentration (C) versus RU plot of CD8-specific binding to K^b-OVA (CD8 over MHC class II I-A^d was the negative control). *c*, The association phase can be fitted on a single exponential model, whereas only the early part (first 12 s) of the dissociation phase (*d*) satisfies such a model. Theoretical curves are shown as smooth lines.

FIG. 2 CD8 enhances the binding of MHC molecules to immobilized TCR molecules as measured by surface plasmon resonance. *A*, Enhancement of binding of L^d-p2Ca complexes to 2C TCR. Binding was tested on immobilized 2C TCR by SPR in the absence and presence of 5 μ M CD8 $\alpha\beta$ heterodimers equilibrated in the mobile phase. *a*, Overlay of 10 μ M L^d-p2Ca with and without CD8; the same concentrations of controls (L^d-MCMV) are also shown. *b*, *c*, Fit of the dissociation phase of L^d-p2Ca from 2C TCR in the absence (*b*) and presence of CD8 (*c*) (*i* and *j* are the two components of the two-dissociation model). *B*, Binding of D^b-M80 complexes to HY TCR in the presence of CD8 $\alpha\beta$, tested over HY surface in the absence (*a*) and in the presence (*b*) of 5 μ M of CD8 $\alpha\beta$.

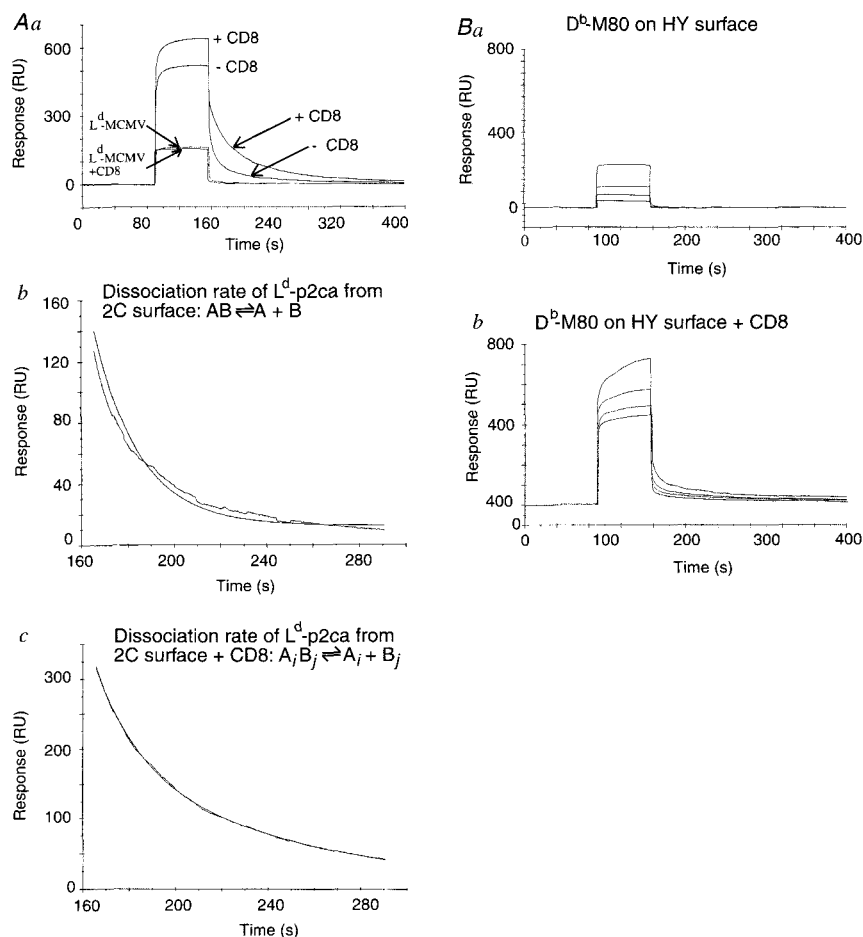
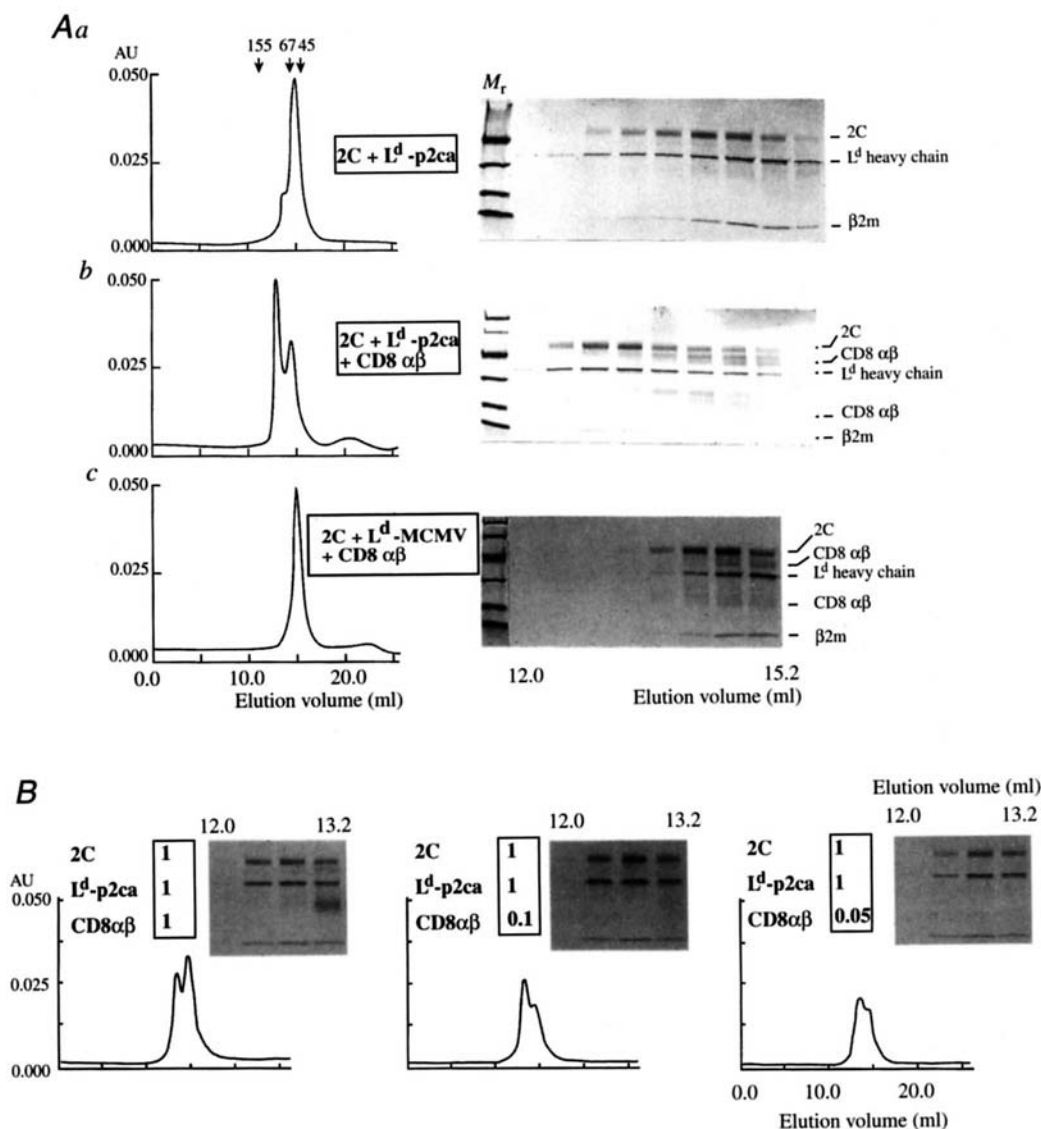


FIG. 3 Isolation of L^d -p2Ca/2C TCR complexes by gel filtration in the presence of CD8 $\alpha\beta$. **A**, Different mixtures (a–c) were analysed by gel filtration and Coomassie R250-stained SDS-PAGE of the peak fractions (gels on the right). The presence of $\alpha\beta$ CD8 dimers generated a greater proportion of TCR/MHC complexes. The higher- M_r and lower- M_r CD8 $\alpha\beta$ represent the disulphide- and non-disulphide-bonded heterodimer. The 2C/ L^d -MCMV/CD8 mixture represents the negative control. β_2m , β_2 -microglobulin; AU, absorbance at 280 nm. **B**, Formation of stable TCR/MHC complexes does not require stoichiometric ratios of CD8. The amount of uncomplexed L^d molecules increased only slightly when CD8 $\alpha\beta$ was present at a 1:20 ratio. Gel-filtration chromatograms and SDS-PAGE of the peak fractions (100K peak) are shown for each experiment. The respective ratios of the 3 molecules are boxed.



actions by SPR on the BIAcore (see 7Methods). The association phase was fitted satisfactorily on a single exponential model and the association constant k_{ass} measured at $3.3 \pm 0.4 \times 10^3 M^{-1} s^{-1}$ (Fig. 1c). The early part of the dissociation phase (Fig. 1d), critical for measuring fast 'off' rates, could be fitted on a single exponential model (dissociation constants, k_{diss} $0.044 \pm 0.005 s^{-1}$), whereas fitting of more extended dissociation periods required a two-exponential model. Note that in this two-exponential dissociation model, the first k_{diss} value was equivalent to the calculated k_{diss} of the early part of the curve, indicating that an accurate early k_{diss} measurement can be extracted. The second, very slow k_{diss} of $5.9 \times 10^{-4} s^{-1}$ corresponded to the baseline shift (4–10% of total bound resonance units (RUs)) and could be fitted on a single exponential model (data not shown). Scatchard calculations yield a K_D (k_{diss}/k_{ass}) of 22 μM for this particular interaction (Fig. 1b). Kinetic measurements of the different CD8/MHC class I (MHC-I) interactions are summarized in Table 1. MHC haplotype and peptide content do not appear substantially to modify the binding to CD8, consistent with the binding of CD8 to conserved (that is, non-polymorphic) MHC residues. Additionally, the binding of the CD8 $\alpha\beta$ and $\alpha\alpha$ forms to MHC class I molecules is very similar, with a slightly higher affinity for the CD8 heterodimer owing to a faster 'on' rate (Table 1). Overall, the MHC-I/CD8 interaction has a moderate affinity (K_D values ranging from 11 to 39 μM), a slow association rate (1.25 to $3.8 \times 10^3 M^{-1} s^{-1}$ of k_{ass}) and a moderate

dissociation rate (0.042 to $0.049 s^{-1}$; half-life, $t_{1/2}$, of 14 to 16 seconds). Further analysis of the SPR data (not shown) and gel filtration (Fig. 3) indicates that both the CD8 $\alpha\alpha$ and CD8 $\alpha\beta$ /MHC interactions are monovalent.

The affinities of TCRs for MHC class I and class II molecules have been shown to range from 45 to 100 μM (refs 1–6). The TCR 2C, however, binds to H-2 L^d with significantly higher affinity^{3,5}; TCR 2C is the clonotypic receptor of a cytotoxic T lymphocyte (CTL) clone educated on K^b molecules but with specificity for the MHC-I alloantigen L^d complexed with a self peptide, p2Ca (sequence: LSPFPFDL)^{13,14}. SPR indicates a measurable interaction of 2C with L^d -p2Ca, but not with L^d bound to a negative control peptide MCMV (sequence: YPHFMPTNL)^{3,5} (Fig. 2). The association and dissociation phases of this interaction could be fitted by a single-site association model with k_{ass} $8.30 \pm 1.45 \times 10^3 M^{-1} s^{-1}$, and k_{diss} $0.027 \pm 0.004 s^{-1}$ ($t_{1/2} = 26 s$) (Figure 2a, b) for a calculated K_D of $3.30 \pm 0.4 \mu M$. This kinetic evaluation of K_D was in agreement with the Scatchard analysis equilibrium measurement of 3 μM (data not shown).

To test the effect of CD8 on the TCR/MHC-I interaction, we repeated the experiment by equilibrating the mobile phases with CD8 before all MHC injections to ensure that the SPR readout reflected only the MHC/TCR interaction. In the presence of 5 μM purified CD8 $\alpha\beta$ dimers, the kinetics of L^d -p2Ca binding to 2C TCR were modified substantially (Fig. 2). The association phase

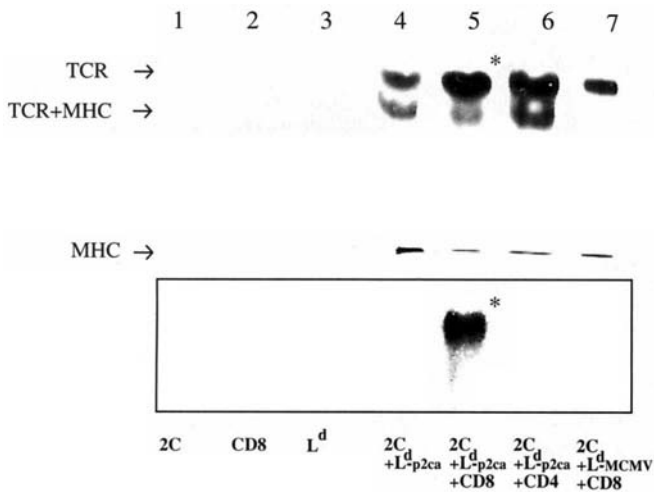


FIG. 4 Isolation of TCR/MHC/CD8 complex on a non-denaturing polyacrylamide gel. CD8 does not enter the gel under basic conditions. In the presence of CD8, the TCR/MHC band, which runs below the TCR band, is replaced by a more intense band running closer to the anode (top) (asterisk, lane 5). Western blot analysis (lower panel) reveals the presence of CD8 in this band (asterisk). The composition of each sample in lanes 1–7 is indicated below the western blot.

could still be fitted to a single-site association model and showed a slight increase in k_{ass} to $1.2 \pm 0.3 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$. The dissociation phase could be fitted satisfactorily to a single-site association model but the best fit of the dissociation phase now showed a parallel dissociation of two complexes, with a fast ‘off’ rate of $0.028 \pm 0.003 \text{ s}^{-1}$ ($t_{1/2} = 25 \text{ s}$), corresponding to the original 2C/ L^d -p2Ca interaction, and a slow ‘off’ rate of $0.0038 \pm 0.0004 \text{ s}^{-1}$ ($t_{1/2} = 184 \text{ s}$), corresponding to the effect of CD8 on the 2C/ L^d -p2Ca interaction (Fig. 24, c). As lower concentrations of L^d -peptide complexes could be fitted on a single-site dissociation model (that is, all complexes slowly dissociating in the presence of CD8), the two-component kinetic model probably indicates that the short duration of the SPR experiments ($<1 \text{ min}$) at the highest L^d -p2Ca concentrations does not allow sufficient time for the CD8 effect to go to completion. Gel filtration confirmed that, when the system is given time to preincubate to equilibrium, this CD8 effect is evident even with substoichiometric ratios of CD8 to TCR–MHC (Fig. 3).

Our affinity measurements of the 2C/ L^d -p2Ca interaction in the absence of CD8 show a K_D value that is almost a log higher than reported for affinity measurements on live $CD8^+$ cells with soluble MHC molecules^{5,7}. In the presence of CD8, our affinity measurements are in good agreement with live $CD8^+$ T-cell measurements^{5,7}.

A low-affinity TCR/MHC-I pair was then examined for any influence of CD8. The interaction of the HY TCR with H-2D^b-M80 (see 7Methods) gave an ‘on’ rate of $6.2 \pm 2.4 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ and an ‘off’ rate of $0.145 \pm 0.02 \text{ s}^{-1}$ ($t_{1/2} = 4.8 \text{ s}$) by SPR. This interaction was dramatically increased in the presence of $CD8\alpha\beta$ dimers (Fig. 2B). The association phase was only slightly altered, with a k_{ass} of $5.1 \pm 0.4 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, but the dissociation phase could be fitted with a two-exponential model with a fast k_{diss} of $0.150 \pm 0.01 \text{ s}^{-1}$ ($t_{1/2} = 6 \text{ s}$), corresponding to the original ‘off’ rate, and a slow k_{diss} of $0.01 \pm 0.005 \text{ s}^{-1}$ ($t_{1/2} = 70 \text{ s}$), accounting for the effect of CD8 (Fig. 2B).

We next tested whether the CD8 effect apparent by SPR would

enable TCR/MHC complexes to be isolated by gel-filtration chromatography, which would also reveal the valency of the complexes by direct measurement of molecular mass. Because of its higher affinity, the 2C TCR/ L^d system was chosen. L^d -p2Ca complexes were incubated to equilibrium in the presence of 2C with or without $CD8\alpha\beta$ heterodimers (Fig. 3). In the presence of the irrelevant L^d -MCMV complex (Fig. 3A, c), the elution profile resolved into a single 45K–55K peak in which 2C, $CD8\alpha\beta$ and L^d -MCMV were all recovered. When 2C was incubated with L^d -p2Ca (Fig. 3A, a), a shoulder of higher M_r (100K) appeared which contained both 2C and L^d -p2Ca, indicating formation of a complex consisting of one TCR and one MHC molecule. The presence of CD4, as a control, did not significantly influence this profile (data not shown). When CD8 was added to 2C and L^d -p2Ca, a large, fully separated peak replaced the 100K shoulder (Fig. 3A, b). SDS–PAGE analysis of gel-filtration column fractions confirmed 2C/ L^d -p2Ca complex formation. In the presence of $CD8\alpha\beta$, almost all L^d molecules were shifted into the additional 100K TCR/MHC complex peak. $CD8\alpha\alpha$ dimers induced similar stabilization (data not shown). Complex formation was surprisingly enhanced for substoichiometric ratios of CD8 to MHC and TCR (Fig. 3B). Only a slight diminution of the stabilization effect resulted when CD8 was present at a 1:20 ratio.

Ternary complexes of CD8, TCR and MHC ($\sim 150\text{K}$) were only detectable by western blotting of the 2C/ L^d -p2Ca/CD8 gel filtration fractions (data not shown), indicative of a transitory, rapidly dissociating TCR/MHC/CD8 complex. This ternary complex was confirmed by non-denaturing (native) gel electrophoresis (Fig. 4). Because of its very basic pI, CD8 does not enter non-denaturing gels under basic conditions, whereas L^d runs close to the front and 2C at an intermediate position. Incubation of 2C with L^d -p2Ca gave an additional band due to TCR/MHC complex formation. In the presence of CD8, but not CD4, this band shifted to a more basic pI. Western blotting using an anti- $CD8\alpha$ antibody confirmed the presence of CD8. An equivalent band was not seen for samples containing an irrelevant peptide, indicating that dimeric CD8/MHC and/or CD8/TCR did not account for the presence of CD8. The high molar concentrations and lack of dilution in the gel assay probably enabled the ternary complex to be detected.

Our results complement previous studies indicating that CD8 adhesion to MHC is influenced by engaging the TCR¹⁵ and that anti-CD8 antibodies can either inhibit or enhance TCR/MHC interactions¹⁶. All these results suggest that CD8 is critical for T-cell recognition, and that CD8 is likely to interact directly with the TCR: a direct interaction of

TABLE 1 Kinetic binding characteristics of CD8 $\alpha\alpha$ homo- and $\alpha\beta$ heterodimers				
CD8	MHC	k_{ass} ($10^3 \times \text{M}^{-1} \text{ s}^{-1}$)	k_{diss} (s^{-1})	K_D (10^{-6} M) ($k_{\text{diss}}/k_{\text{ass}}$)
CD8 $\alpha\alpha$	K ^b -OVA	1.4 ± 0.2	0.044 ± 0.002	30.4 ± 3.8
CD8 $\alpha\alpha$	K ^b -VSV	1.2 ± 0.4	0.049 ± 0.002	39.3 ± 5.0
CD8 $\alpha\beta$	K ^b -OVA	3.3 ± 0.4	0.044 ± 0.005	14.0 ± 0.2
CD8 $\alpha\beta$	K ^b -VSV	3.8 ± 0.6	0.045 ± 0.003	11.8 ± 1.3
CD8 $\alpha\beta$	L ^d -p2Ca	3.6 ± 0.4	0.042 ± 0.005	11.2 ± 0.8
CD8 $\alpha\beta$	D ^b -M80	3.4 ± 0.3	0.048 ± 0.005	14.1 ± 0.4

Each measurement is the average of 3 independent experiments.

the TCR with CD4 and CD8 has been proposed before^{17–20}.

The identification of a ternary TCR/MHC/CD8 interaction and the stabilization of the TCR/MHC complex by CD8 can be explained in three ways: (1) by a statistical model of rapid serial engagement of CD8 which results in transient stabilization of the TCR/MHC complex, followed by disengagement of CD8 and the quantitative appearance of long-lived TCR/MHC complexes; (2) by a CD8-induced conformational change which results in an energetically more favourable complex formation; (3) a combination of (1) and (2), where there is still rapid turnover of CD8 engagement and disengagement, but CD8 binding is accompanied by structural perturbations that increase stability. Conformational rearrangement occurs in other immune recognition events, as demonstrated in crystal structures of Fab fragments of antibodies in response to antigen binding²¹. CD8 may induce structural changes in the binding site that enhance complementarity at the interface between TCR and MHC.

The requirement for an induced fit in the CD8/MHC/TCR interaction could vary from case to case, as seen in Fab-antigen interactions²¹. The relation between the *in vitro* phenomenon described here and the situation *in vivo* needs to be determined in order to evaluate the role of co-capping, local density and lateral motions, for example, of the different receptors in the physiological environment. □

Methods

cDNA constructs, protein expression and purification. All cDNAs were modified by PCR by using overhang primers retaining modified sequences and cloning sites. MHC heavy chains were truncated after codon 274 (Trp), CD8 α after codon 156 (Ser), and CD8 β after codon 146 (Lys). For TCR α - and β -chains, the cytoplasmic tails were retained in the soluble constructs. For the 2C (V α 3-J α 58C α ; V β 8.2D β 2-J β 2.4C β 2) TCR α -chain, the cDNA was interrupted after codon 221 (Asn) and extended by the sequence corresponding to residues 244–248 of its cytoplasmic tail (RLWSS). 2C TCR β -chain was interrupted at codon 246 (Gln) and extended by the sequence corresponding to residues 278–283 of the β -chain cytoplasmic tail (VKKKNS). Soluble HY TCR (V α 1, V β 8.1; O. Kanagawa, unpublished result) was produced using the same strategy. Soluble four-domain murine CD4 was engineered by deletion of the cytoplasmic and transmembrane regions of the full-length cDNA²² (interruption at codon 372). The tailored cDNAs were subcloned between the EcoRI and SalI sites of the polylinker of pRMHa3, a metallothionein promoter-drive vector¹¹. The different constructs were transfected into SC2 fly cells by calcium phosphate precipitation together with pUChsneo, a neomycin-resistance gene¹¹. Stable cells lines were derived by G418 selection and kept under selection in Schneider's medium (Gibco BRL) containing 500 μ g of G418 per ml. Protein was induced over a period of 3 d by adding 700 μ M copper sulphate to cells expended in serum-free medium (Insect X-Press, Biowhittaker). Supernatants were collected, concentrated and dialysed by ultrafiltration and filtered. MHC molecules were purified by Ni-agarose (Ni²⁺-NTA-agarose; Qiagen) step-gradient affinity chromatography (8 mM and 20 mM imidazole, pH 7.2, washes; 150 mM imidazole elution) and ion-exchange chromatography (Mono-Q in 50 mM Tris, pH 8.0, 0–500 mM NaCl gradient; Pharmacia). TCR molecules were purified from supernatants by Ni-agarose affinity chromatography, gel filtration (Sephacryl S-100; Pharmacia) in PBS, pH 7.4, and hydrophobic-interaction chromatography (phenyl Superose (Pharmacia); loading in 1M (NH₄)₂SO₄/50 mM phosphate buffer, pH 7.5, elution by 0–100% gradient of 50 mM phosphate

buffer). The native conformation of purified TCR molecules was verified by antibody binding using SPR; anti-idiotypic 1B2 (ref. 23) or anti-V β 8 F23.1 (ref. 24), anti-C α H28-710 (ref. 25), anti-C β H57-597 (Pharmingen) were used in the case of 2C TCR. The anti-idiotypic MR2.6 antibody Fab (O. Kanagawa, unpublished results) was used for HY TCR. The MHC molecules K^b and L^d were stored at high concentrations (15 mg ml⁻¹), with a slight molar excess of peptide. Before use, MHC-peptide preparations were checked by native gel electrophoresis and isoelectric focusing to ensure that complex formation was complete. CD8 molecules were purified by Ni-agarose chromatography, followed by gel filtration and cation exchange chromatography (Mono-S (Pharmacia), with 50 mM Tris, pH 8.0, 0–0.5 M NaCl gradient). The CD8 homo- and heterodimers were a mixture of disulphide-bonded and non-disulphide-bonded dimers. Purified I-A^b molecules were prepared as before⁵. Soluble CD4 was purified by Ni-agarose affinity chromatography, gel filtration and hydrophobic interaction chromatography (C.S. and L.T., unpublished data).

Surface plasmon resonance. Experiments were run on a BIAcore 2000 machine (Biosensor, NJ) at 25 °C. Purified molecules were immobilized on a CM5 sensor chip at pH 5.5 in 10 mM acetate buffer by classic amine coupling. For Fig. 1, 1,800 RU of K^b-OVA complexes were immobilized, where >90% of the immobilized protein was active, as shown by injecting a conformation-specific antibody, Y3 (ref. 26). Blank surfaces were made by ethanolamine deactivation of the activated surface. For Fig. 1, different ligands were injected at 20 μ l min⁻¹ at 30, 15, 7.5, 3.75 and 1.875 μ M (top to bottom traces) in PBS over the K^b-OVA surface. For Fig. 2, 1,100 RU of purified 2C TCR and 1,600 RU of HY were immobilized. L^d complexes were injected at 20, 15, 5 and 2.5 μ M at 20 μ l min⁻¹; D^b complexes were injected at 40, 20, 10 and 5 μ M. For injection in the presence of CD8, the flow cell and tubing were equilibrated with 5 μ M purified CD8 in PBS using the 'Prime' command. At high ligand concentrations, the baseline can drift after injection (4–10% of total bound RU) owing to matrix effects and the high protein concentrations required for low-affinity interactions. SPR data were analysed with BIAevaluation 2.1 software. Model curve fitting was chosen from a ln(R₀/R) against t plot of the original data. ($R = R_0 e^{-k_d(t-t_0)}$), where R₀ is the response at the start of dissociation). The whole dissociation phase as well as its early part (first 10–30 s) and late part (last 2 min) were fitted. Scatchard plots were derived from the resonance unit (RU)/concentration (C) versus RU plot, and linear regression (Cricket Graph software) of specific binding.

Gel filtration, SDS and native polyacrylamide gel electrophoresis, and western blotting. Gel filtration was done in PBS on a Superdex-200 column at 0.4 ml min⁻¹. Equilibrium conditions were determined by incubating 2C TCR with L^d-p2Ca for 1 h, 4 h or overnight at 37 °C. Each mixture was incubated for 4 h at 37 °C and analysed by gel filtration. Peak fractions were analysed by SDS-PAGE gels (10–20% gradient gels; Biorad). Low-molecular-mass markers were used (Biorad; 14.4K, 21K, 31K, 45K, 67K, 97K). For Fig. 3A, L^d molecules were in a 2-fold excess over 2C molecules, but were equimolar in all other experiments. 2C TCR was 50 μ M in all experiments, CD8 25 μ M, and CD4 50 μ M. Non-denaturing gels (homogenous, 7.5% acrylamide pH 8.8) were run on a PHAST system (Pharmacia). For western blotting, SDS-PAGE and non-denaturing native gels were transferred to a PVDF membrane by semi-dry blotting. CD8 was detected by using either a rabbit anti-serum raised against recombinant CD8 dimers or the anti-CD8 monoclonal antibody 53–6.7 (Pharmingen). Specific binding was revealed by a peroxidase-coupled goat anti-rabbit and chemiluminescence.

Peptides. Peptides were synthesized by t-BOC chemistry on an Applied Biosystems instrument and purified on a C18 reverse-phase column (Rainin Instruments). For the HY TCR, which is specific for an unknown peptide derived from the H-Y male antigen bound to the MHC class I-D^b molecule, peptide mimotopes were derived by M. Gavin and M. Bevan using an HY TCR-bearing cell line OH2.4 (O. Kanagawa). M80 (sequence FAGHNLDLI) is a pure agonist of OH2.4 lytic activity. D^b-SEV (FAPGNYPAL) complexes were used as a negative control. VSV peptide sequence is RGVVYQGL.

Received 18 June; accepted 30 October 1996.

- Matsui, K. et al. *Science* **254**, 1788–1791 (1991).
- Weber, S., Trautwein, A., Oliveri, F., Gerhard, W. & Karjalainen, K. *Nature* **356**, 793–795 (1992).
- Corr, M. et al. *Science* **265**, 946–949 (1994).
- Matsui, K., Boniface, J. J., Steffner, P., Reay, P. A. & Davis, M. M. *Proc. Natl Acad. Sci. USA* **91**, 12862–12866 (1994).
- Sykulev, Y. et al. *Immunity* **1**, 15–22 (1994).
- Scott, C. A., Garcia, K. C., Carbone, F. R., Wilson, I. A. & Teyton, L. J. *Exp. Med.* **183**, 2087–2095 (1996).
- Sykulev, Y. et al. *Proc. Natl Acad. Sci. USA* **91**, 11487–11491 (1994).
- Doyle, C. & Strominger, J. L. *Nature* **330**, 256–259 (1987).
- Potter, T. A., Rajan, T. V., Dick, R. F. & Bluestone, J. A. *Nature* **337**, 73–75 (1989).
- Salter, R. D. et al. *Nature* **345**, 41–46 (1990).
- Matsumura, M., Saito, Y., Jackson, M. R., Song, E. S. & Peterson, P. A. *J. Biol. Chem.* **267**, 23589–23595 (1992).
- Van der Merve, P. A., Brown, M. H., Davis, S. J. & Barclay, A. N. *EMBO J.* **12**, 4945–4954 (1993).
- Udaka, K., Tsomides, T. J. & Eisen, H. N. *Cell* **69**, 989–998 (1992).
- Udaka, K., Tsomides, T. J., Walden, P., Fukusen, N. & Eisen, H. N. *Proc. Natl Acad. Sci. USA* **90**, 11271–11276 (1993).
- O'Rourke, A. M., Rogers, J. & Mescher, M. F. *Nature* **346**, 187–189 (1990).
- Luescher, I. F. et al. *Nature* **373**, 353–356 (1995).

- Rojo, J. M., Saizawa, K. & Janeway, C. A. Jr *Proc. Natl Acad. Sci. USA* **86**, 3311–3315 (1989).
- Gallagher, P. F., Fazekas de St Groth, B. & Miller, J. F. A. P. *Proc. Natl Acad. Sci. USA* **86**, 10044–10048 (1989).
- Hoo, W. S. & Kranz, D. M. *J. Immunol.* **150**, 4331–4337 (1993).
- Vignali, D. A. A., Carson, R. T., Chang, B., Mittler, R. S. & Strominger, J. L. *J. Exp. Med.* **183**, 2097–2108 (1996).
- Stanfield, R. L., Takimoto-Kamimura, M., Rini, J. M., Profy, A. T. & Wilson, I. A. *Structure* **1**, 83–93 (1993).
- Tourville, B., Gorman, S. D., Field, S. H., Hunkapiller, T. & Barnes, J. R. *Science* **234**, 610–614 (1986).
- Kranz, D. M., Tonesgawa, S. & Eisen, H. N. *Proc. Natl Acad. Sci. USA* **81**, 7922–7926 (1984).
- Staerz, U. D. et al. *J. Immunol.* **134**, 3994–4000 (1985).
- Becker, M. L. B. et al. *Cell* **58**, 911–921 (1989).
- Ajtkumar, P. et al. *Cell* **58**, 911–921 (1989).

ACKNOWLEDGEMENTS. We thank L. Hatlen for protein production; R. Teasdale for anti-CD8 serum; D. Kranz and O. Kanagawa for cDNAs and antibodies for the 2C TCR and HY TCR, respectively; M. Gavin and M. Bevan for identifying mimotopes for the HY TCR; R. Boismenu for cloning; and the R. W. Johnson P. R. I. peptide laboratory for peptide synthesis and protein sequencing. I.A.W. is supported by the NIH. K.C.G. is a recipient of an NIH postdoctoral training grant.

CORRESPONDENCE and requests for materials should be addressed to L.T.

Import of carrier proteins into the mitochondrial inner membrane mediated by Tim22

Christian Sirrenberg, Matthias F. Bauer,
Bernard Guiard*, Walter Neupert & Michael Brunner

Institut für Physiologische Chemie der Universität München,
Goethestrasse 33, 80336 München, Germany
* Centre de Genétique Moléculaire CNRS, Université Pierre et Marie Curie,
91190 Gif-sur-Yvette, France

TRANSLOCATION of mitochondrial preproteins across the inner membrane is facilitated by the TIM machinery¹⁻⁸. Tim23 binds to matrix targeting signals and initiates membrane potential-dependent import⁸. Tim23 and Tim17 are constituents of a translocation channel across the inner membrane⁶. Tim44 is associated with this channel at the matrix side⁶, and Tim44 recruits mitochondrial Hsp70 and its co-chaperone Mge1, which drive protein translocation into the matrix using ATP as an energy source⁹⁻¹⁴. Tim22 is a new component of the import machinery of mitochondria, which shares sequence similarity with both Tim23 and Tim17. Here we report that Tim22 is required for the import of proteins of the mitochondrial ADP/ATP carrier (AAC) family into the inner membrane. Members of the yeast AAC family are synthesized without matrix targeting signals^{15,16}. Tim22 is in an assembly of high relative molecular mass that is distinct from the Tim23–Tim17 complex⁶. Import of proteins of the AAC family is independent of Tim23, and import of matrix targeting signals containing preproteins is independent of Tim22.

An open reading frame on *Saccharomyces cerevisiae* chromosome IV (D0884) has the potential to encode a transmembrane protein with a relative molecular mass (M_r) of 21.8K, henceforth referred to as Tim22. The second half of the predicted Tim22 sequence is similar to the membrane-integrated portions of Tim23 and Tim17, components of the translocase of the mitochondrial inner membrane²⁻⁸ (Fig. 1). Antibodies directed against the carboxy terminus of Tim22 detected a protein of M_r 22K in yeast extracts, indicating that Tim22 is expressed in *S. cerevisiae*. Upon subcellular fractionation, Tim22 was enriched in the mitochondrial fraction but was virtually absent from the microsomal fraction (Fig. 2a). It was inaccessible to proteases added to intact mitochondria, but was accessible to proteases added to mitoplasts (data not shown). Submitochondrial particles were generated by sonication and separated by sucrose gradient sedimentation. Tim22 cofractionated with inner-membrane particles, and was absent in fractions containing outer-membrane vesicles (Fig. 2b). Tim22 was not extracted from membranes by carbonate at pH11 (data not shown). These observations demonstrate that Tim22 is an integral inner-membrane protein of the mitochondria.

Mitochondria were made soluble using digitonin, and gel filtration was performed. Tim22 was in an assembly of high M_r , which eluted from the sizing column in a fraction corresponding to 300K (Fig. 2c); it did not coelute with the Tim23–Tim17 complex⁶. Tim22 was not precipitated from mitochondria made soluble using detergent by antibodies against Tim23 or Tim17, and anti-Tim22 IgG did not precipitate Tim23 or Tim17 (data not shown). These results suggest that Tim22 is present in a complex of high M_r , which is distinct from the Tim23–Tim17 complex.

Tim22 was disrupted in the diploid yeast strain MB2 (ref. 1) by insertion of the HIS3 marker, and cells were allowed to sporulate. Tetrad analysis yielded only two viable spores (Fig. 3a), which were *his*⁻ and expressed Tim22 (data not shown), indicating that Tim22 is essential for the viability of *S. cerevisiae*. We used the haploid strain 344, which cannot metabolize galactose as a carbon source¹⁷, to construct Tim22(Gal10). In this strain the Gal10

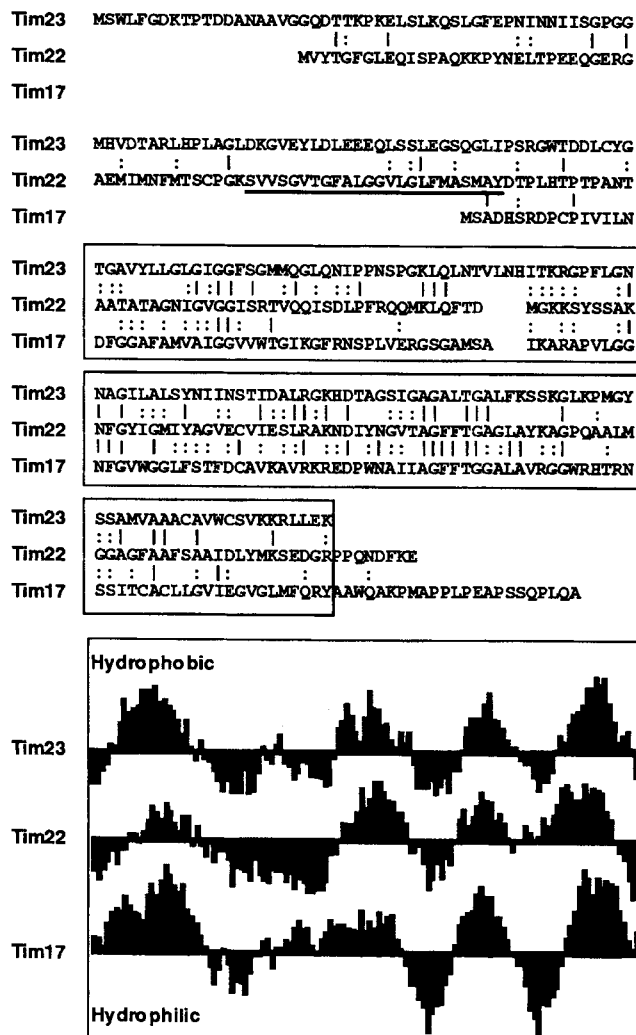


FIG. 1 Tim22 shares sequence similarity with Tim23 and Tim17. Alignment of the sequences of Tim23, Tim22 and Tim17. Identical (connecting line) and similar residues (colons) between sequences (ref. 25) are indicated. Hydropathy profiles are shown for the boxed region. A potential membrane-spanning segment in the N-terminal portion of Tim22 is underlined.

promoter was integrated into the chromosome in front of Tim22, allowing regulated expression of Tim22 (Fig. 3b). In the presence of galactose, Tim22(Gal10) grew like the parental 344 strain (Fig. 3c), but shifting to galactose-free medium led to a reduced growth rate, and growth virtually stopped after 4 days.

Mitochondria were prepared from 334 cells (wild type), from Tim22(Gal10) cells grown in the presence of galactose (Tim22⁺), and from Tim22(Gal10) cells two days (Tim22[↓]) and three days (Tim22^{↓↓}) after the shift to galactose-free medium. They were analysed by western blotting for various mitochondrial proteins (Fig. 3d). Tim22 was not detected in mitochondria after the shift to galactose-free medium, but was present in wild-type mitochondria and those from cells grown in the presence of galactose. Levels of the ADP/ATP carrier (AAC)¹⁵ and the phosphate carrier (PiC)¹⁶ were slightly reduced in Tim22[↓] mitochondria, and both proteins were virtually absent in Tim22^{↓↓} mitochondria. Otherwise mitochondria from cells in galactose-free medium did not show an altered protein composition, as they contained normal levels of marker proteins for the outer membrane, the intermembrane space, the matrix, and the inner membrane. Furthermore, cytochrome spectra were not altered in mitochondria from cells grown in galactose-free medium, and both trans-

lated mitochondrially encoded proteins (data not shown). This suggests that Tim22 is specifically involved in biogenesis or turnover of members of the mitochondrial AAC family.

Energized mitochondria from wild-type cells and from cells two days after the shift to galactose-free medium were incubated with radiolabelled AAC, PiC, precursor Su9(1-79)-dihydrofolate reductase¹⁸ (pSu9(1-79)-DHFR) and precursor cytochrome *b*₂Δ19(1-167)-DHFR_{K5} (pb₂Δ19(1-167)-DHFR_{K5}) (Fig. 4a). Import of the preproteins was subsequently assayed by acquisition of protease resistance in the mitochondria. To monitor insertion of AAC and PiC into the inner membrane and translocation of pSu9(1-79)-DHFR and pb₂Δ19(1-167)-DHFR_{K5} across the inner membrane into the matrix, the outer membrane was opened by osmotic swelling of the mitochondria, and the resulting mitoplasts were treated with proteinase K. AAC and PiC were imported into wild-type mitochondria and were inserted into the inner membrane. In contrast, AAC and PiC were not imported into mitochondria from cells that had been grown on galactose-free medium for two days. The matrix-targeted preproteins pSu9(1-79)-DHFR and pb₂Δ19(1-167)-DHFR_{K5} were efficiently imported into both types of mitochondria, suggesting that Tim22 facilitates the import of AAC and PiC into the inner membrane but is not required for the import of preproteins into the matrix.

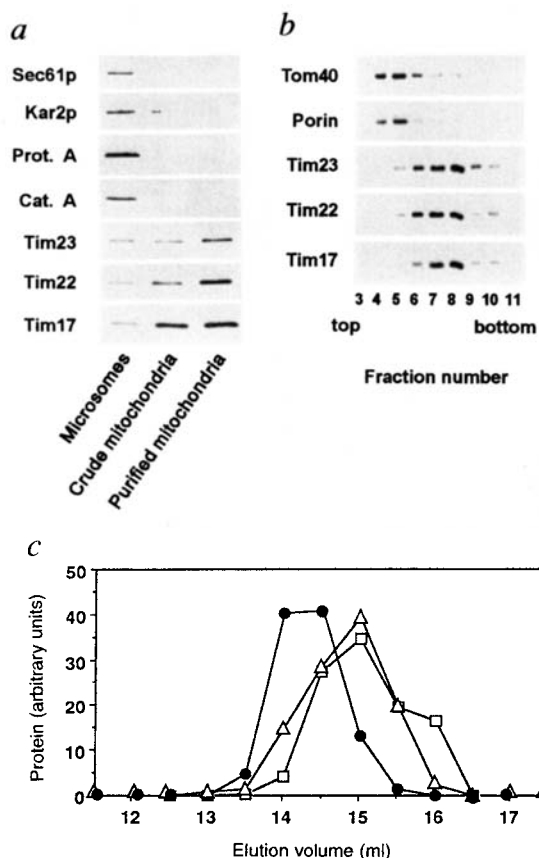


FIG. 2 Tim22 is a protein of the mitochondrial inner membrane. *a*, Microsomes, a crude mitochondrial fraction, and purified mitochondria were assayed by immunoblotting for Sec61 and Kar2 (endoplasmic reticulum), protease A (Prot. A) and catalase A (Cat. A) (vacuoles and peroxisomes, respectively), Tim23 and Tim17 (mitochondria), and Tim22. *b*, Tim22 is localized in the inner membrane. Submitochondrial particles were separated on a sucrose gradient²⁶. Fractions were analysed by immunoblotting using antibodies against the indicated proteins. *c*, Tim22 (circles) is not in a complex with Tim23 (squares) and Tim17 (triangles). Mitochondria were solubilized with digitonin and applied onto a Superose-6 column⁶. Fractions (0.5 ml) were analysed by immunoblotting.

In the absence of a membrane potential, AAC accumulates as a productive translocation intermediate (stage 3) on its pathway to the inner membrane¹⁹. In this situation, AAC is resistant to low levels of externally added protease; it is in association with components of the TOM complex partly translocated across the outer membrane²⁰. Mitochondria from wild-type cells and from cells two days after the shift to galactose-free medium were preincubated either with respiratory chain substrates to increase the membrane potential, or with valinomycin/KCl to dissipate the membrane potential. Radiolabelled AAC was then imported, and mitochondria were subsequently treated with 30 μg ml⁻¹ proteinase K to monitor accumulation of AAC at stage 3, and with 200 μg ml⁻¹ proteinase K to assay for completely imported species

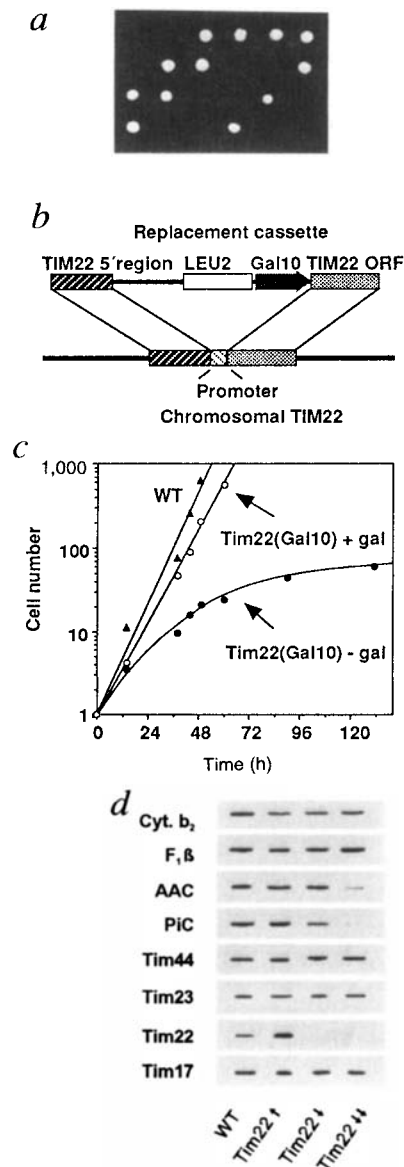


FIG. 3 Tim22 is required for biogenesis of carrier proteins. *a*, Dissection of tetrads after sporulation of a TIM22 heterozygous diploid (a/α MB2-TIM22/tim22::HIS3). *b*, Construction of Tim22(Gal10) by replacement of the chromosomal TIM22 promoter by the Gal10 promoter. The 'replacement cassette' is outlined schematically. *c*, Downregulation of Tim22 affects cell growth. Tim22(Gal10) cells were grown in the presence or absence of galactose. The cell number at time 0 was set equal to 1. *d*, Tim22 is required to maintain carrier proteins. Mitochondria were prepared from wild-type (WT) and from Tim22(Gal10) cells grown in lactate medium with galactose (Tim22↑), without galactose for two days (Tim22↓), and without galactose for three days (Tim22↓↓). Mitochondria were assayed for the indicated antigens. Cyt. *b*₂, cytochrome *b*₂; F₁β, β-subunit of F₁-ATPase.

(Fig. 4b). AAC was imported in wild-type mitochondria in the presence of membrane potential and accumulated at stage 3 in the absence of membrane potential. In mitochondria from cells two days after the switch to galactose-free medium, AAC accumulated at stage 3 in the presence and absence of membrane potential. Thus Tim22 is required for translocation of AAC from the TOM complex to the TIM machinery in a reaction that requires membrane potential.

AAC was imported into wild-type and Tim22^{-/-} mitochondria in the presence of membrane potential, and into wild-type mitochondria in the absence of membrane potential. Chemical cross-linking was performed with disuccinimidyl glutarate (DSG) or *m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester (MBS). Mitochondria were subsequently solubilized using detergent and were immunoprecipitated with antibodies against Tim22 (Fig. 4c). A radiolabelled adduct of $M_r \sim 55K$ was precipitated from

wild-type mitochondria when crosslinking with DSG or with MBS was done with a membrane potential present. The adduct corresponded in size to a crosslink of AAC with Tim22, but was not detected in wild-type mitochondria in the absence of membrane potential ($\Delta\Psi$), nor in mitochondria from cells two days after switching to galactose-free medium in the presence of membrane potential. This demonstrates that Tim22 is in contact with AAC in the presence of membrane potential, but not when its import is arrested at stage 3.

Taken together, our data demonstrate that Tim22 is required at or before the membrane potential-dependent step for import into the inner membrane of AAC and PiC, both of which are members of the mitochondrial AAC family. Tim22 does not seem to be required for the import of preproteins targeted to the matrix space.

To relate the Tim22-dependent import pathway to protein translocation by means of the Tim23–Tim17 complex⁶ and mitochondrial Hsp70, mitochondria were prepared which were compromised in the function of Tim23 or mitochondrial Hsp70. A mutant yeast strain was constructed that encodes Tim23(fs), a variant of Tim23 with a translational frameshift after methionine 178. Yeast cells expressing Tim23(fs) are viable but the mitochondria contain reduced levels of Tim23(fs) (Fig. 5a). Mitochondria were prepared from the *sscl-3* strain, which encodes a mutant form of mitochondrial Hsp70 that is irreversibly inactivated by

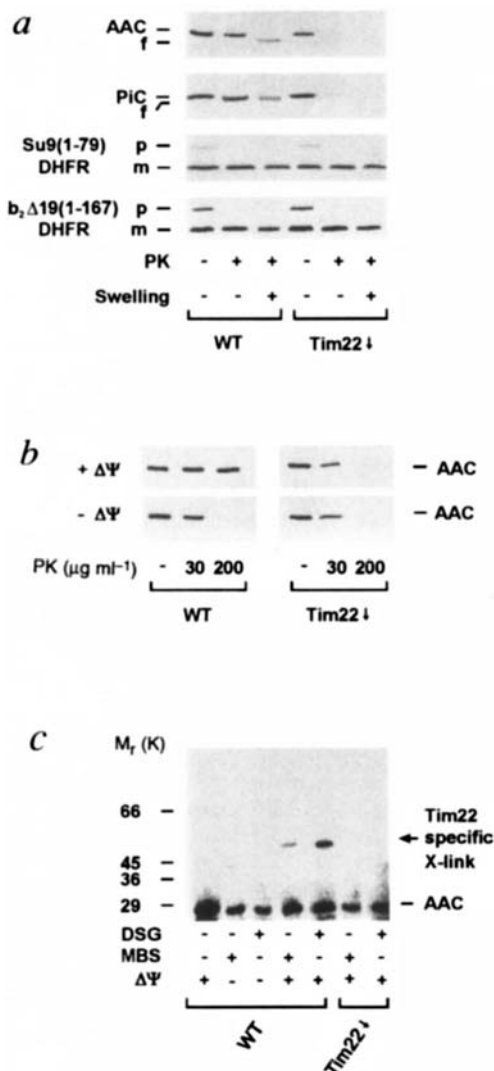


FIG. 4 Tim22 facilitates the import of carrier proteins. *a*, Mitochondria from cells grown without galactose for two days (Tim22⁻) are defective in import of AAC and PiC. Preproteins were imported into wild-type (WT) and Tim22⁻ mitochondria¹⁸. Portions were left untreated, or treated with proteinase K (PK), or converted to mitoplasts in the presence of PK¹⁸. f, protease resistant fragment; p, precursor; m, mature form. *b*, Import of AAC is arrested at stage 3 in Tim22⁻ mitochondria. $\Delta\Psi$, membrane potential. *c*, Tim22 crosslinked to AAC during import. AAC was imported at 10 °C in the presence of DSG (200 μ M) or MBS (200 μ M). Mitochondria were solubilized and immunoprecipitated with anti-Tim22 IgG.

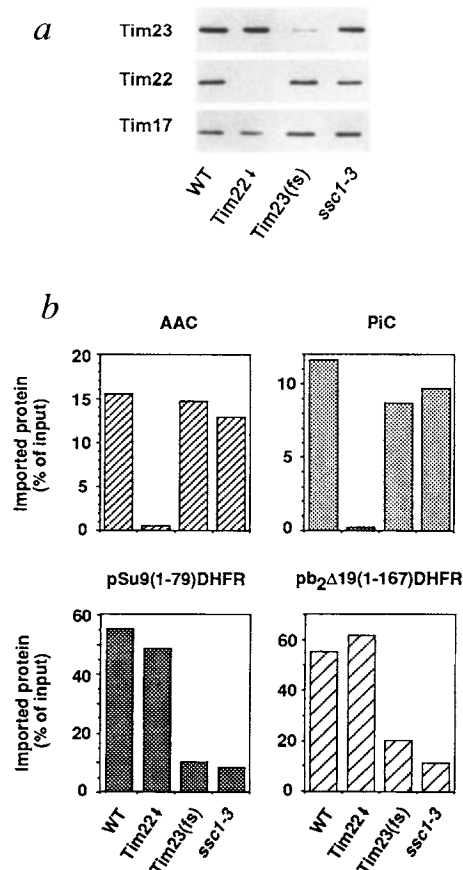


FIG. 5 Tim22 and Tim23—mitochondrial Hsp70 are required for import of different subsets of preproteins. *a*, Mitochondria prepared from wild-type (WT), from Tim22(Gal10) cells grown without galactose for two days (Tim22^Δ), Tim23(fs), and *ssc1-3* were assayed for Tim23, Tim22 and Tim17. *b*, AAC, Pic, pSu9(1–79)–DHFR, pb Δ 19(1–167)–DHFR were imported into the different mitochondria. Samples were treated with PK, analysed by SDS–PAGE and autoradiography, and quantified with a phosphor imaging system¹⁴.

incubation of the mitochondria at 37 °C (refs 14, 21).

AAC, PiC, pSu9(1–79)–DHFR and pb₂Δ19(1–167)–DHFR_{K5} were incubated with energized mitochondria (wild-type, Tim22^Δ, Tim23(fs) and *sscl-3*), and import of the preproteins was measured (Fig. 5b). The preproteins were efficiently imported into wild-type mitochondria. Mitochondria from the galactose-free cells imported pSu9(1–79)–DHFR and pb₂Δ19(1–167)–DHFR_{K5}, but did not import AAC and PiC. In contrast, import of pSu9(1–79)–DHFR and pb₂Δ19(1–167)–DHFR_{K5} was drastically reduced in Tim23(fs) and *sscl-3* mitochondria, but import of AAC and PiC was as efficient as in wild-type mitochondria. These observations indicate that Tim22 acts on a pathway parallel to those of Tim23 and mitochondrial Hsp70.

In summary, mitochondria contain distinct pathways for targeting of preproteins to the matrix space and for insertion of precursors of the mitochondrial AAC family into the inner membrane. Preproteins with a matrix targeting signal are imported by means of the Tim23–Tim17 complex⁶ driven by mitochondrial Hsp70–Tim44–Mge1 (ref. 14). Tim23 acts as a receptor for matrix targeting signals, and triggers membrane potential-dependent translocation of the presequence across the inner membrane⁸. On the matrix side, Tim44, mitochondrial Hsp70 and Mge1 then act together to drive complete translocation of the polypeptide chain, with ATP as an energy source^{6,12–14}. Yeast mitochondria contain about 35 members of the AAC family, which constitute a substantial fraction of the total mitochondrial protein; in bovine mitochondria, for instance, AAC and PiC constitute about 15% of the inner membrane proteins^{22,23}. Members of the yeast AAC family do not have an amino-terminal matrix targeting signal. We have shown that import into the inner membrane of two members of the AAC family requires Tim22 function but is independent of the Tim23–Tim17 complex and mitochondrial Hsp70. In view of the structural similarities of the various members of the AAC family, we anticipate that other members also use the Tim22 pathway for import. Tim22 and Tim23–Tim17 translocases can function independently of each other. However, it remains possible that, in the case of membrane proteins that contain both a matrix targeting signal and internal membrane insertion signals, Tim22 might cooperate with Tim23–Tim17. □

Methods

Disruption of Tim22. TIM22 DNA was amplified by the polymerase chain reaction (PCR) using the upstream primer 5'-CCCCGAATTCATCGAGCTTCACTGAA-3' and the downstream primer 5'-CCCGATATCGACAGTCTCTAACGCTC-3'. The PCR fragment was digested with *EcoRI* and *SalI* and cloned into pGEM4, yielding pGEM4–TIM22c. A blunt-ended DNA fragment containing the *HIS3* gene was inserted into the *StuI* site of TIM22, yielding pGEM4–tim22::HIS3. The plasmid was digested with *EcoRI* and *SalI*, and used to transform the diploid yeast strain MB2 (ref. 1). *HIS*⁺ colonies were selected and allowed to sporulate. Tetrads were dissected as described²⁴.

Replacement of the TIM22 promoter. A *BamHI*–*XbaI* fragment encoding TIM22 was amplified by PCR and cloned into Yep51–Tim44myc¹³ yielding Yep51–Tim22myc. A PCR fragment corresponding to the 5' untranslated region of TIM22 (–368 to –68 with respect to the first nucleotide of the open reading frame) was then subcloned into pGEM4. The fragment was excised with *ScaI* and ligated into Yep51–Tim22myc, replacing the 2μ yeast replication origin on the *PvuII*–*HpaI* fragment. The resulting 'replacement cassette' plasmid was linearized with *SstI* and used to transform the haploid yeast strain 334 (ref. 17). *LEU*⁺ colonies were analysed for the ability to express chromosomal TIM22 under galactose control. The resulting strain is referred to as Tim22(Gal10).

Immunoprecipitation of crosslinked proteins. Crosslinking was performed for 30 min at 10 °C. The crosslinker was quenched with 100 mM glycine. Mitochondria were reisolated and solubilized with 0.5% SDS. Samples were diluted to a final concentration of 0.025% SDS and immunoprecipitation with affinity-purified anti-Tim22 IgG was performed.

Construction of Tim23(fs). Yep51–Tim23 (ref. 6) was linearized with *NcoI*, which cleaves at the codon for Met 178 of Tim23. The ends were filled in with Klenow enzyme and religated yielding Yep51–Tim23(fs). TIM23 heterozygous diploid (a/α MB2–Tim23/tim23::HIS3) was transformed with Yep51–TIM23(fs) and transformants were allowed to sporulate. After dissection of tetrads *LEU*⁺, *HIS*⁺ spores were selected.

Received 9 October; accepted 11 November 1996.

- Maarse, A. C., Blom, J., Grivell, L. A. & Meijer, M. *EMBO J.* **11**, 3619–3628 (1992).
- Emtage, J. L. T. & Jensen, R. E. *J. Cell Biol.* **122**, 1003–1012 (1993).
- Dekker, P. J. T. et al. *FEBS Lett.* **331**, 66–70 (1993).
- Ryan, K. R., Menold, M. M., Garrett, S. & Jensen, R. E. *Mol. Biol. Cell* **5**, 529–538 (1994).
- Maarse, A. C., Blom, J., Keil, P., Pfanner, N. & Meijer, M. *FEBS Lett.* **349**, 215–221 (1994).
- Berthold, J. et al. *Cell* **81**, 1085–1093 (1995).
- Blom, J., Dekker, J. T. & Meijer, M. *Eur. J. Biochem.* **232**, 309–314 (1996).
- Bauer, M. F., Sirennberg, C., Neupert, W. & Brunner, M. *Cell* **87**, 33–41 (1996).
- Kang, P.-J. et al. *Nature* **348**, 137–143 (1990).
- Laloraya, S., Gambill, B. D. & Craig, E. A. *Proc. Natl Acad. Sci. USA* **91**, 6481–6485 (1994).
- Kronidou, N. G. et al. *Proc. Natl Acad. Sci. USA* **91**, 12818–12822 (1994).
- Rassow, J. et al. *J. Cell Biol.* **127**, 1547–1556 (1994).
- Schneider, H. C. et al. *Nature* **371**, 768–774 (1994).
- Schneider, H. C., Westermann, B., Neupert, W. & Brunner, M. *EMBO J.* **15**, 5796–5803 (1996).
- Zimmermann, R., Paluch, U., Sprinzl, M. & Neupert, W. *Eur. J. Biochem.* **99**, 247–252 (1979).
- Zara, V. et al. *Eur. J. Biochem.* **198**, 405–410 (1991).
- Hovland, P., Flick, J., Johnston, M. & Scalafani, R. A. *Gene* **83**, 57–64 (1989).
- Ungermann, C., Guillard, B., Neupert, W. & Cyr, D. M. *EMBO J.* **15**, 735–744 (1996).
- Pfanner, N. & Neupert, W. *J. Biol. Chem.* **262**, 7528–7536 (1987).
- Söllner, T. et al. *Nature* **355**, 84–87 (1992).
- Gambill, B. D. et al. *J. Cell Biol.* **123**, 109–117 (1993).
- Klingenberg, M., Riccio, P. & Aquila, H. *Biochim. Biophys. Acta* **503**, 193–210 (1978).
- Kolbe, H. V. J., Costello, D., Wong, A., Lu, R. C. & Wohlrab, H. *J. Biol. Chem.* **259**, 9115–9120 (1984).
- Rowley, N. et al. *Cell* **77**, 249–259 (1994).
- Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. *Atlas of Protein Sequence and Structure* Vol 5 (suppl. 3) (ed. Dayhoff, M. O.) 342–352 (National Biomedical Research Foundation, Silver Spring, MD, USA).
- Pon, L., Moll, T., Vestweber, D., Marshallsay, B. & Schatz, G. J. *Cell Biol.* **109**, 2603–2616 (1989).

ACKNOWLEDGEMENTS. We thank A. Zollner for help with database searches and, C. Kotthoff and C. Zobywalski for technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 184 (Teilprojekt B12), the Human Frontiers of Science Program, the Fonds der Chemischen Industrie, and the Münchener Medizinische Wochenschrift (to M.F.B. and M.B.).

CORRESPONDENCE and requests for materials should be addressed to W.N. (e-mail: neupert@bio.med.uni-muenchen.de). The sequence of Tim22 has been deposited in PIR database, accession number S67776.

Translocation arrest of an intramitochondrial sorting signal next to Tim11 at the inner-membrane import site

Kostas Tokatlidis, Tina Junne, Suzette Moes, Gottfried Schatz, Benjamin S. Glick* & Nafsika Kronidou

Biozentrum, University of Basel, CH-4056 Basel, Switzerland
* Cummings Life Science Center, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA

THE import of proteins from the cytosol into the mitochondrial matrix involves the concerted action of two separate import systems: the TOM system in the outer membrane, and the TIM system in the inner membrane^{1,2}. Here we report that the inner-membrane system also sorts proteins to the intermembrane space. Some intermembrane-space proteins, such as cytochromes *b*₂ and *c*₁, are synthesized with a complex pre-sequence consisting of a positively charged matrix targeting signal followed by an uncharged sequence that acts as sorting signal for the intermembrane space³. We show that this sorting signal can be efficiently crosslinked to an inner-membrane protein of relative molecular mass 11K after the mature part of the precursor has been sorted to the intermembrane space. The 11K protein, which we term Tim11, is a component of the protein import system in the inner membrane.

We began by fusing the cytochrome *b*₂ presequence to mouse dihydrofolate reductase (DHFR), generating a hybrid precursor targeted to the intermembrane space⁴. This precursor was then

modified to contain a unique cysteine between the uncharged stretch and the cleavage site for the inner-membrane protease that removes the sorting signal⁵. The precursor was then imported into isolated yeast mitochondria under conditions in which the DHFR moiety was sorted to the intermembrane space, but cleaved only slowly from the sorting signal. The sorting signal was then crosslinked to an adjacent protein by disulphide crosslinking. Finally, the crosslinked protein was purified and characterized.

When the cysteine-containing (G76C) cytochrome *b*₂-DHFR precursor (pb₂-DHFR) or a cysteine-free variant were imported into isolated yeast mitochondria, both precursors reached a protease-inaccessible location in intact mitochondria (Fig. 1, lanes 2, 6). Import required an electrochemical potential across the inner membrane (lanes 4, 10). Because the import reaction was brief, and cleavage by the inner-membrane protease is slow for precursors that lack the folded haem-binding domain of cytochrome *b*₂ (ref. 6), more than 50% of the imported molecules were cleaved only by the matrix-processing peptidase. These once-cleaved molecules accumulated as the translocation intermediate^{7,8}, which was anchored by the cytochrome *b*₂ sorting signal to the outer face of the inner membrane.

If a thiol group is activated with 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB), it can react with a second thiol group to form a stable disulphide bridge⁹. The cysteine-containing pb₂-DHFR precursor (G76C) was activated with DTNB and imported into mitochondria. More than 80% of the imported molecules were recovered as a 36K species, indicating that a crosslink had formed (Fig. 1). When the experiment was repeated with the cysteine-free

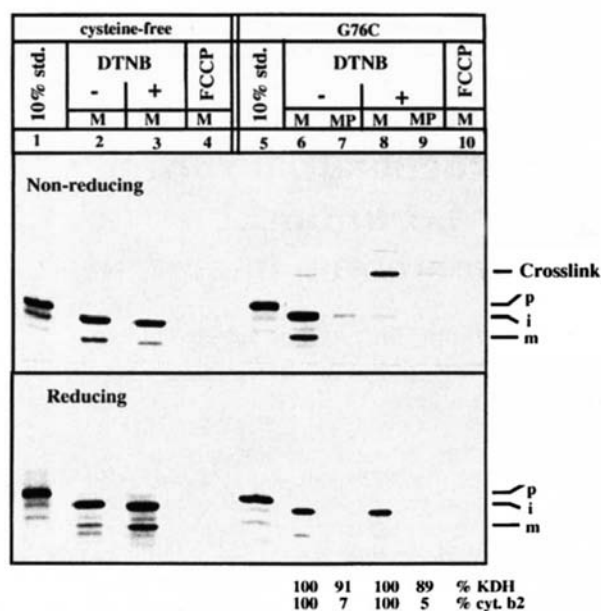


FIG. 1 The pb₂-DHFR translocation intermediate crosslinks efficiently to an 11K membrane protein. The cysteine-containing pb₂-DHFR (G76C) was activated with DTNB. Top, analysis by non-reducing SDS-PAGE; bottom, analysis of the same samples by reducing SDS-PAGE. M, mitochondria treated with trypsin after import; MP, mitoplasts generated in the presence of proteinase K. Lanes 1 and 5, 10% of the precursor added to each sample; 2-4, cysteine-free pb₂-DHFR variant; 6-10, G76C variant of pb₂-DHFR; 2 and 6, untreated precursors; 3 and 8, DTNB-derivatized precursors; 4 and 10, mitochondria de-energized by the uncoupler carbonyl cyanide *p*-(trifluoromethoxy)phenylhydrazine (FCCP); 7, import of the untreated G76C precursor followed by generation of mitoplasts; and 9, as lane 7 but using the DTNB-derivatized precursor. Mitoplast formation was confirmed by immunoblotting for α -ketoglutarate dehydrogenase (KDH) and cytochrome *b*₂ (cyt. b₂) as markers for the matrix and intermembrane space, respectively; p, precursor form; i, translocation intermediate; m, mature form. The position of the crosslinked product is indicated.

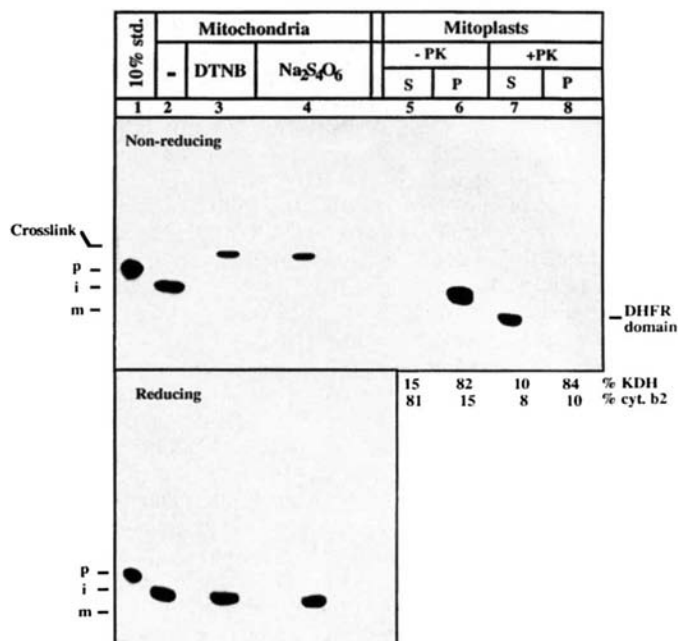


FIG. 2 Crosslinking to the 11K protein can occur after complete import of pb₂-DHFR to the intermembrane space. After import of the cysteine-containing pb₂-DHFR (lane 1, 10% standard), mitochondria were treated with trypsin and divided into four samples; one sample (lane 2) was left untreated; a second sample (lane 4) was crosslinked by addition of Na₂S₂O₆; and the two remaining samples were converted to mitoplasts in the absence (lanes 5 and 6) or presence (lanes 7 and 8) of proteinase K (PK), and then separated by centrifugation into supernatant (S) and pellet (P) fractions. In a separate control reaction (lane 3), pb₂-DHFR was activated with DTNB before import. Top, analysis by non-reducing SDS-PAGE; bottom, analysis of the same samples by reducing SDS-PAGE. The protease-resistant DHFR fragment is indicated.

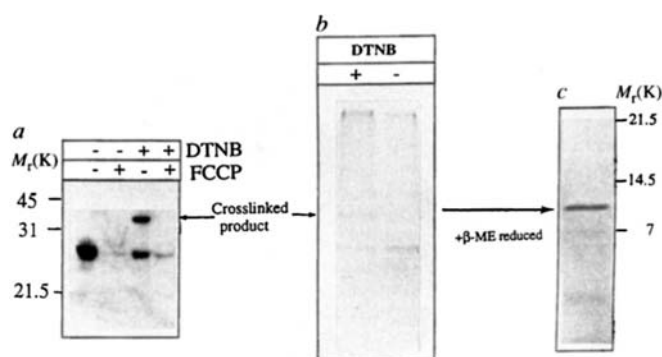


FIG. 3 Isolation of the crosslinked 11K protein. a, Western blot. For each sample shown, 0.2 mg of purified cysteine-containing pb₂-DHFR was either left untreated or activated with DTNB, then imported into 10 mg of purified mitochondria in the presence or absence of FCCP. The translocation intermediate and crosslinked product were detected using SDS-PAGE followed by immunoblotting with an anti-DHFR antibody. b, Coomassie stain. Import reactions were performed as in a, but with 100 mg of mitochondria and 2 mg of precursor. The crosslinked translocation intermediate was purified on Ni²⁺-NTA-agarose beads and subjected to non-reducing SDS-PAGE followed by brief staining with Coomassie blue. For microsequencing, the material of ten such experiments was pooled. c, Silver stain. The crosslinked material from b was excised from the gel and reduced with β -mercaptoethanol (β -ME), then electrophoresed on a second gel and visualized by silver staining.

pb₂-DHFR precursor, the 36K species was not formed. This species represented the intermediate form of pb₂-DHFR linked by a disulphide bond to an unlabelled protein, as demonstrated by SDS-PAGE under reducing conditions, in which the radioactivity originally present in the 36K species was quantitatively recovered as the intermediate form (Fig. 1, bottom). The intermediate form of pb₂-DHFR (*M_r* 25K) had therefore been crosslinked to a protein of *M_r* ~ 11K.

When the mitochondrial outer membrane was disrupted selectively to generate mitoplasts, both the non-crosslinked intermediate form of pb₂-DHFR (Fig. 1) and the crosslinked 36K species became accessible to added protease. Thus introduction of a cysteine or activation with DTNB do not compromise the sorting function of the cytochrome *b₂* presequence, and the crosslinked translocation intermediate is anchored to the outer face of the inner membrane. As the crosslinked species was largely resistant to extraction under alkaline conditions (data not shown), the 11K protein seemed to be tightly associated with the inner membrane.

The crosslink can form after the precursor has been sorted to the intermembrane space (Fig. 2). When we first imported the underivatized pb₂-DHFR precursor into mitochondria, and then generated the crosslink by adding the oxidant sodium tetrathionate¹⁰, the 36K crosslinked product was again formed with an efficiency above 80%, and could be reductively cleaved to the intermediate form of pb₂-DHFR. The high crosslinking efficiency suggests that we are examining the main (and probably the only) pathway followed by the cytochrome *b₂* sorting signal.

The pb₂-DHFR translocation intermediate seems to be anchored to the inner membrane with its mature DHFR portion entirely within the intermembrane space (Fig. 1), so the DHFR moiety should be able to fold. To test this prediction, mitochondria that had imported pb₂-DHFR without crosslinking were converted to mitoplasts in the presence of methotrexate. In the absence of protease, the imported intermediate form of pb₂-DHFR remained associated with the mitoplast pellet (Fig. 2). When proteinase K was added to the mitoplasts, the DHFR moiety was released into the supernatant as a large fragment. Generation of this characteristic fragment¹¹ confirmed that the DHFR moiety had folded in the intermembrane space, and that pb₂-DHFR had completed its sorting reaction before the crosslinking reaction.

The experiment shown in Fig. 2 was performed under standard conditions in which the mitochondrial matrix contains high levels of ATP¹². Depleting the matrix of ATP slowed, but did not block, the correct import of pb₂-DHFR, in agreement with earlier work using a related precursor⁶. The efficiency of tetrathionate-mediated crosslinking to the 11K protein was unaffected by depleting ATP in the matrix.

The unusually efficient 'zero length' crosslinking suggested that the crosslinked mitochondrial protein is closely apposed to the membrane-anchored sorting signal, and that the arrested pb₂-DHFR intermediate might serve as an affinity reagent for purifying the 11K protein. However, the experiments shown in Figs 1 and 2 were done with only a few nanograms of radiolabelled precursor. Import and crosslinking reactions were therefore scaled up by a factor of approximately one million (Fig. 3). The crosslinked protein could now be detected by protein staining as a distinct 11K band. After microsequencing two tryptic peptides, we found in the yeast genome database an open reading frame of 96 codons that encoded a protein of *M_r* 10,867 which contained both peptide sequences (Fig. 4a). This predicted 11K protein also contained a single cysteine residue at position 28, a putative transmembrane region at its amino terminus, and a hydrophilic carboxy-terminal portion. The unique cysteine residue was located immediately after the hydrophobic region. Because the protein proved to be a component of the protein import complex of the mitochondrial inner membrane, we named it Tim11 (ref. 2).

Disruption of the *Tim11* gene in haploid cells slowed growth on all carbon sources, and diminished mitochondrial respiration and cytochrome oxidase activity by 50% (data not shown). Tim11 is

a 1 M S T V N V L R Y S A L G L G L F F G F 20
21 R N D M I L K C N A K K K E E Q A Q Y E 40
41 E K L K L V E E A K K E Y A K L H P V V 60
61 T P K D V P A N A S F N L E D P N I D F 80
81 E R V I L N A V E S L K E A S T 96

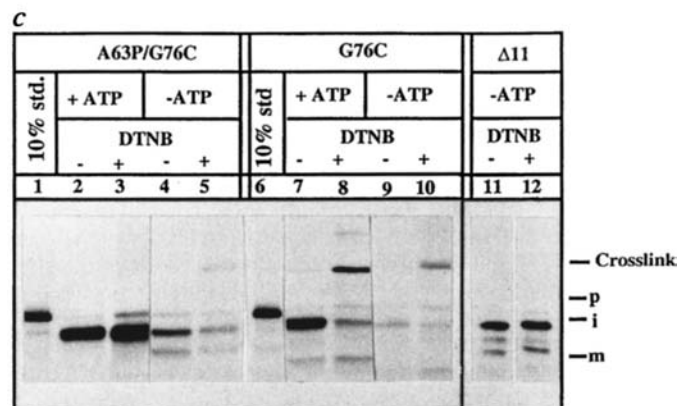
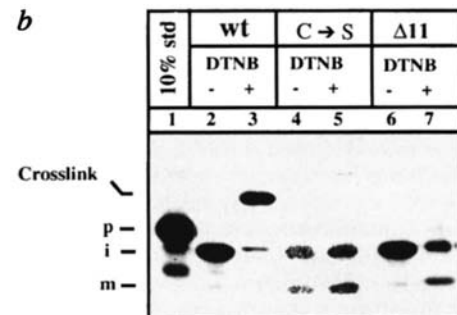


FIG. 4 The crosslinked 11K protein is a component of the protein import system in the inner membrane. a, Predicted amino-acid sequence of Tim11. The unique cysteine at position 28 is boxed, and the two microsequenced tryptic peptides are underlined. Dotted underlining marks positions in the tryptic peptides at which the sequencing cycle failed to identify an amino acid. b, Tim11 is the protein that becomes crosslinked to the pb₂-DHFR translocation intermediate. The cysteine-containing pb₂-DHFR (lane 1) was either left untreated (-) or activated with DTNB (+), then imported into either wild-type (wt) mitochondria (lanes 2 and 3), mitochondria in which cysteine 28 of Tim11 had been changed to serine (C → S; lanes 4 and 5), or mitochondria from cells lacking Tim11 (Δ11; lanes 6 and 7). c, A matrix-targeted precursor arrested in transit through the inner membrane can be crosslinked to Tim11. Lanes 1–5, the matrix-targeted A63P variant of the cysteine-containing pb₂-DHFR (A63P/G76C); 6–10, the cysteine-containing pb₂-DHFR precursor (G76C); 1 and 6, 10% of the amount of precursor added to each import assay; 2, 3, 7 and 8 (+ATP), import into fully energized mitochondria without (-) or with (+) prior activation of the precursor by DTNB; 4, 5, 9 and 10 (-ATP), import into ATP-depleted mitochondria; and 11 and 12, import of the A63P/G76C precursor into ATP-depleted Tim11-free mitochondria.

thus not essential, but is required for optimal mitochondrial function.

The experiment of Fig. 4b confirmed that Tim11 is the protein that was crosslinked to pb₂-DHFR. DTNB-activated pb₂-DHFR was imported into mitochondria that either lacked Tim11 or that contained a Tim11 variant bearing a serine instead of the unique cysteine. Cells expressing the cysteine-less mutant of Tim11 were indistinguishable from wild-type cells in their growth rate and mitochondrial function (data not shown), and neither of the two mitochondria yielded the crosslinked product. The crosslinking reaction thus involves formation of a disulphide bond between

cysteine 76 of pb₂-DHFR and cysteine 28 of Tim11.

Previous studies had suggested that the pb₂-DHFR translocation intermediate is anchored in the protein import system of the inner membrane^{7,13}. If Tim11 is part of this system, it should also be in the immediate vicinity of a matrix-targeted precursor trapped during its passage across the inner membrane. Matrix-targeted precursors can be trapped in the inner membrane system by depleting ATP in the matrix^{12,14}. The resulting 'ATP-depletion intermediates', the mature part of which is located in the intermembrane space¹⁴⁻¹⁶, can be crosslinked specifically to components of the inner-membrane import system¹⁷⁻¹⁹.

We converted pb₂-DHFR to a matrix-targeted precursor by mutating alanine 63 to proline, thereby inactivating the sorting signal²⁰. When this matrix-targeted pb₂-DHFR variant (A63P/G76C) was activated with DTNB and imported into ATP-depleted mitochondria, it became crosslinked to Tim11 (Fig. 4c, lane 5). Crosslinking was again very efficient, approaching 40% of the translocation intermediate, but was lower than with the 'wild type' intermediate on its way to the intermembrane space. We suspect that the specific sorting signal of the wild-type precursor positions the precursor's sulphhydryl group more precisely next to Tim11 than does a general translocation arrest caused by inactivating the matrix-located import motor. When the mitochondria are solubilized by using a variety of detergents, Tim11 does not co-immunoprecipitate with Tim23 (K.T. *et al.*, unpublished observation). Nevertheless, our results (Fig. 4) establish Tim11 as a component of the protein import system in the inner membrane.

When the matrix targeted pb₂-DHFR variant was presented to ATP-sufficient mitochondria, it was imported completely into the matrix without forming a crosslink. In contrast, 'wild type' pb₂-DHFR became crosslinked to Tim11 both in ATP-depleted and in ATP-sufficient mitochondria. When the matrix-targeted pb₂-DHFR variant was imported into ATP-depleted mitochondria lacking Tim11 (Fig. 4, lanes 11 and 12), it did not yield any crosslink; when it was crosslinked to Tim11 in ATP-depleted wild-type mitochondria, it could be cleaved quantitatively from Tim11 and chased into the matrix by using dithiothreitol plus ATP (data not shown). Because Tim11 can be crosslinked to a true translocation intermediate of a matrix-targeted precursor, it must be in the immediate vicinity or part of the protein import site in the inner membrane.

These findings have a bearing on the long-standing question of how proteins such as cytochromes b₂ and c₁ are sorted to the mitochondrial intermembrane space. The two suggested pathways differ critically in how the sorting signal is decoded: the 'stop-transfer' pathway^{7,13} proposes that this is done by the inner-membrane import system, whereas the 'conservative sorting' pathway^{3,21} invokes a separate re-export system. Our results demonstrate that after pb₂-DHFR has been imported to the intermembrane space, its sorting signal is still closely attached to a component of the inner-membrane import system. Tim11 has thus enabled us to show that the sorting signal of a precursor that has already been transported to the intermembrane space is in the same, or a very similar, environment in the inner membrane as a precursor on its way to the matrix. □

Methods

Constructs. The pSP65-based plasmid encoding pb₂(1-167)-DHFR (ref. 8) was modified using polymerase chain reaction (PCR)-mediated gene fusion to encode the first 84 residues of the cytochrome b₂ precursor fused to residue 2 of a mouse DHFR variant with serine at position 7. Cysteine 14 was then converted to alanine, generating pSM18. Unique cysteine codons (TGC) were introduced into pSM18 to generate pSM19 (S52C), pSM20 (A64C), pSM21 (A69C), pSM22 (G59C) and pSM23 (G76C). Plasmid pKT4, which encodes the matrix-targeted A63P/G76G variant of pb₂-DHFR, was constructed by mutating the alanine-63 codon in pSM23 to CCC (proline).

Protein import. Precursors were synthesized in a reticulocyte lysate⁷, denatured by 8 M urea, treated with 0.5 mM DTNB for 30 min at room temperature (where indicated), and diluted 20-fold into standard import buffer containing

yeast mitochondria, apyrase, carboxyatractylide and NADH¹². Import was allowed to proceed for 10 min at 20 °C. Other procedures were as described previously⁷. Immunoblots were developed with [¹²⁵I]protein A and quantified using a β-imager (Biospace, France), taking the amount of a given protein present in the intact mitochondria as 100%.

Crosslinking. Crosslinking with 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) was performed by activating cysteine 76 in pb₂-DHFR or its variants with DTNB, and importing the derivatized precursor into mitochondria. For tetrathionate treatment, the mitochondria were allowed to import the underivatized precursor, and were washed, resuspended in import buffer containing 0.1 mM freshly prepared Na₂S₄O₆, and left for 60 min on ice. Higher concentrations of Na₂S₄O₆ inhibited disulphide formation (data not shown), probably because they derivatized both pb₂-DHFR and Tim11 (refs 10, 22).

Isolation of Tim11. The modified cytochrome b₂ presequence was fused to hexahistidine-tagged DHFR by replacing a portion of DHFR in plasmid pQE16 (Qiagen), from the start codon to the SacI site, with the corresponding portion of the pb₂-DHFR gene from pSM23. To stabilize the fusion protein against proteolysis, leucine 2 was changed to valine, yielding plasmid pKT1. The His-tagged pb₂-DHFR was overexpressed in *Escherichia coli* strain SG13009 (Qiagen), where it formed inclusion bodies. These were washed and solubilized²³ in 8 M urea, 20 mM imidazole-HCl, pH 8.0, and passed over a G-25 column yielding a 70% pure precursor. We performed 200 separate import reactions with 20 mg of DTNB-derivatized pb₂-DHFR and 1 g of gradient-purified mitochondria²⁴. The import mixtures were supplemented with glycerokinase and oligomycin to deplete external and intramitochondrial ATP¹². After import, the mitochondria were treated with proteinase K, washed and solubilized at 5 mg ml⁻¹ for 15 min on ice in 1% Triton X-100, 8 M urea, 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole-HCl, pH 8.0, 1 mM phenylmethylsulphonyl fluoride. After centrifugation at 100,000g for 15 min, 10-ml portions of the supernatant were incubated for 1 h at 4 °C with 0.25 ml Ni²⁺-NTA-agarose beads in solubilization buffer. The beads were washed twice, and the bound protein was eluted with 2% SDS, 100 mM EDTA, 50 mM PIPES-NaOH, pH 7.4, precipitated by methanol/chloroform²⁵ suspended in SDS-PAGE sample buffer, and loaded on a non-reducing 10% Tricine-SDS gel²⁶. After the gel was stained with Coomassie blue, the crosslinked product was excised, incubated overnight at room temperature in sample buffer containing β-mercaptoethanol, and electrophoresed on a 16.5% Tricine-SDS gel. The 11k band (approximately 1 μg) was subjected to trypsin digestion and microsequencing.

Cloning and disruption of the TIM11 gene. The previously unidentified open reading frame on *Saccharomyces cerevisiae* chromosome IV (nucleotides 1,112,039 to 1,112,329) was amplified by PCR from yeast genomic DNA and disrupted at the SalI site with the *kanMX2* module²⁷. The TIM11(C28S) allele was generated by replacing codon 28 (TGT) with TCT, and was expressed from plasmid YCplac33 in a Tim11-less yeast strain.

Received 26 July; accepted 11 November 1996.

- Pfanner, N. & Meijer, M. *Curr. Biol.* **5**, 132-135 (1995).
- Pfanner, N. *et al. Trends Biochem. Sci.* **21**, 51-52 (1996).
- Glick, B. S., Beasley, E. M. & Schatz, G. *Trends Biochem. Sci.* **17**, 453-459 (1992).
- van Loon, A. P. G. M., Brändli, A. W. & Schatz, G. *Cell* **44**, 801-812 (1986).
- Nunnari, J., Fox, T. & Walter, P. *Science* **262**, 1997-2004 (1995).
- Glick, B. S., Wachter, C., Reid, G. A. & Schatz, G. *Protein Sci.* **2**, 1901-1917 (1993).
- Glick, B. S. *et al. Cell* **69**, 809-822 (1992).
- Hartl, F.-U., Ostermann, J., Guillard, B. & Neupert, W. *Cell* **51**, 1027-1037 (1987).
- Habeeb, A. F. S. A. *Methods Enzymol.* **25**, 457-464 (1972).
- Phil, A. & Lange, R. J. *Biol. Chem.* **237**, 1356-1362 (1962).
- Eilers, M. & Schatz, G. *Nature* **322**, 228-232 (1986).
- Wachter, C., Schatz, G. & Glick, B. S. *Mol. Biol. Cell* **5**, 465-474 (1994).
- Gärtner, F., Bömer, U., Guillard, B. & Pfanner, N. *EMBO J.* **14**, 1349-1359 (1995).
- Hwang, S. T., Wachter, C. & Schatz, G. *J. Biol. Chem.* **266**, 21083-21089 (1991).
- Wachter, C., Schatz, G. & Glick, B. S. *EMBO J.* **11**, 4787-4794 (1992).
- Jascur, T., Goldenberg, D. P., Vestweber, D. & Schatz, G. *J. Biol. Chem.* **267**, 13636-13641 (1992).
- Scherer, P. E., Manning-Krieg, U. C., Jenö, P., Schatz, G. & Horst, M. *Proc. Natl Acad. Sci. USA* **89**, 11930-11934 (1992).
- Ryan, K. R. & Jensen, R. E. *J. Biol. Chem.* **268**, 23743-23746 (1993).
- Kübrich, M. *et al. FEBS Lett.* **349**, 222-228 (1994).
- Beasley, E. M., Müller, S. & Schatz, G. *EMBO J.* **12**, 2303-2311 (1993).
- Gruhler, A., Ono, H., Guillard, B., Neupert, W. & Stuart, R. A. *EMBO J.* **14**, 1349-1359 (1995).
- Kaufmann, S. H. & Shaper, J. H. *Exp. Cell Res.* **155**, 477-495 (1984).
- Tokatidis, K. *et al. FEBS Lett.* **282**, 205-208 (1991).
- Glick, B. S. & Pon, L. *Methods Enzymol.* **260**, 213-223 (1995).
- Wessel, D. & Glügge, U. I. *Anal. Biochem.* **138**, 141-143 (1984).
- Schägger, H. & von Jagow, G. *Anal. Biochem.* **166**, 368-379 (1987).
- Wach, A., Brachat, A., Pöhlmann, R. & Philippsen, P. *Yeast* **10**, 1791-1808 (1994).

ACKNOWLEDGEMENTS. We thank P. Jenö and P. James for microsequencing, and H. Brüttsch and W. Oppliger for technical assistance. This work was supported by grants from the Swiss National Science Foundation and the Human Capital and Mobility Program of the European Union (G.S.), the Human Frontiers Science Program Organization (K.T. and N.K.), and the Roche Research Foundation (K.T.).

CORRESPONDENCE and requests for materials should be addressed to K.T. (e-mail: tokatidis@ubaclu.unibas.ch).

The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast

Francesco De Rubertis*, David Kadosh†, Sandra Henchoz*, Daniel Pauli*, Gunter Reuter‡, Kevin Struhl† & Pierre Spierer*

* Department of Zoology and Animal Biology, University of Geneva, 30 quai E.-Ansermet, CH-1211 Geneva 4, Switzerland

† Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Institut für Genetik, Martin-Luther-Universität, D-06108 Halle/S, Germany

BOTH position-effect variegation (PEV)^{1,2} in *Drosophila* and telomeric position-effect in yeast (TPE)^{3–5} result from the mosaic inactivation of genes relocated next to a block of centromeric heterochromatin or next to telomeres. In many aspects, these phenomena are analogous to other epigenetic silencing mechanisms, such as the control of homeotic gene clusters⁶, X-chromosome inactivation⁷ and imprinting in mammals⁸, and mating-type control in yeast⁵. Dominant mutations that suppress or enhance PEV are thought to encode either chromatin proteins or factors that directly affect chromatin structure¹. We have identified an insertional mutation in *Drosophila* that enhances PEV and reduces transcription of the gene in the eye–antenna imaginal disc. The gene corresponds to that encoding the transcriptional regulator RPD3 in yeast^{9,10}, and to a human histone deacetylase¹¹. In yeast, *RPD3*-deletion strains show enhanced TPE, suggesting a conserved role of the histone deacetylase RPD3 in counteracting genomic silencing. This function of RPD3, which is in contrast to the general correlation between histone acetylation and increased

transcription, might be due to a specialized chromatin structure at silenced loci.

In the *w^{m4h}* inversion of *Drosophila melanogaster*, the *white* gene is relocated next to a breakpoint within centromeric heterochromatin, and can thus be permanently inactivated in some cells by spreading of the adjacent condensed heterochromatin^{1,2} (Fig. 1). The resulting mottled eye-colour phenotype (Fig. 1a) is sensitive to the dose of several unlinked loci, known as suppressors and enhancers of position-effect variegation. These modifiers respectively decrease and increase the frequency of clones in which the gene is inactive¹. The eye of a fly carrying a homozygous viable P-transposon insertion at 64 BC, on the third chromosome, is shown in Fig. 1b; the insertion enhances the variegation of *white*, which seems to be inactive in almost all ommatidia. In 5–10% of the eyes, patches of ommatidia show the disorganization characteristic of the phenotype of the *roughest* mutation, a locus near *white*, but more distant from heterochromatin in the rearrangement. This suggests that the enhancement of variegation allows the silencing to spread past the *white* locus, and to inactivate the *roughest* locus in a fraction of the cells (Fig. 1c, d). The mutation is homozygous viable, and the strong phenotype seems similar in homozygotes and heterozygotes.

In order to link the enhancer phenotype and the P-element insertion, excisions of the transposon were induced and found fully to revert the enhancer phenotype. The locus was cloned by plasmid rescue, and a single transcript of 2.2 kilobase (kb) long was detected on northern blots of embryonic, larval and adult RNA in a 20-kb genomic interval surrounding the insertion point. The P-element is inserted 1.8 kb from the 5' end of the transcript, in the putative regulatory region (Fig. 2). We were unable to find convincing differences in the developmental transcription profile between wild-type and mutant total RNA. However, *in situ* detection of transcripts on whole-mount larval imaginal discs revealed a significant reduction of transcript level in eye–antenna discs of homozygous *rp3* mutant flies (Fig. 2), whereas other discs from the same animals showed no clear difference, and eye–antenna discs exhibited normal staining for an unrelated gene (*dE2F*; data not shown). This result indicates that enhanced PEV results from the loss of RPD3 function in the phenotypically relevant tissue.

Complementary DNA clones were isolated from libraries of embryonic RNA¹², and a clone with an insert of 2,172 nucleotides (a length similar to that of the 2.2-kb transcript on northern blots) was sequenced. An open reading frame encoding 520 amino acids is preceded by a 5' untranslated leader 223 bases long, and followed by a 3' untranslated sequence 389 bases long. Database analysis reveals that the transcript encodes a new gene in *Drosophila* that strongly resembles the human histone deacetylase RPD3, which has also been cloned in yeast, worm and frog^{9–11} (Fig. 3).

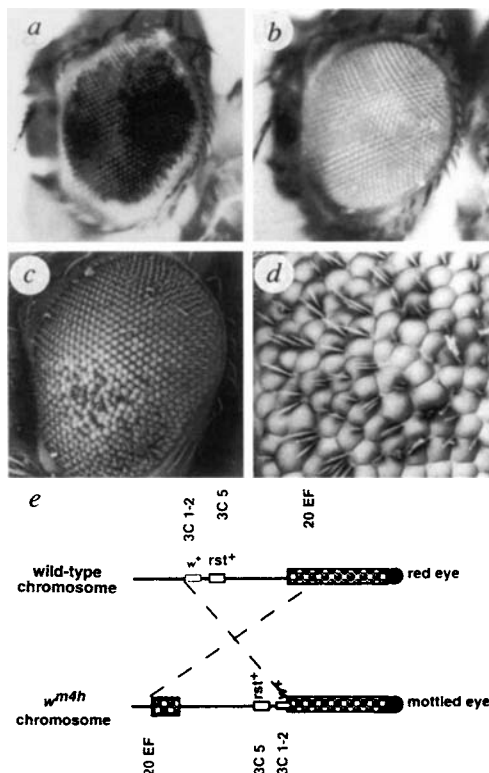


FIG. 1 Effect of an insertional mutation at 64 BC on PEV. a, b, Heads of males carrying the X-chromosome inversion *w^{m4h}* (a), and carrying the same inversion together with the heterozygous mutation *Ev15-1* (b). c, Scanning electron micrograph of the head of a fly of the same genotype as in b; d, enlargement of the *roughest* domain. e, Schematic representation of the *w^{m4h}* inversion.

The high degree of similarity prompted us to test whether a mutation of yeast *RPD3* also affects position-effect variegation. In *Saccharomyces cerevisiae*, genes positioned up to 4.9 kb proximal to a telomere can be transcriptionally silenced, a phenomenon called telomeric position-effect (TPE)³⁻⁵. The closer a gene is located to the telomere, the greater is the silencing¹³. To measure TPE in yeast, we used isogenic wild-type *RPD3* and $\Delta rpd3$ strains that contain a *URA3* gene positioned 2.0, 3.5 and 6.5 kb from the right end telomere of chromosome V¹³. *URA3* expression was monitored by growth on medium containing 5-fluoro-orotic acid (FOA), a drug that is toxic to cells expressing this gene. The $\Delta rpd3$ strains containing the *URA3* gene positioned 2.0 and 3.5 kb from the telomere grew significantly better than did wild-type *RPD3* strains on FOA medium, demonstrating that an *rpd3* null allele increases telomeric silencing at these distances (Fig. 4).

Our results indicate that *RPD3* has an analogous function with respect to genomic silencing in both budding yeast and *Drosophila*. In both organisms, loss of *RPD3* function results in increased genomic silencing, which decreases with the distance from heterochromatin. Although we cannot exclude the possibility that *RPD3* indirectly affects genomic silencing (for example, by repressing the expression of a positive regulator of chromatin condensation), the striking similarity of *rpd3* mutant phenotypes in yeast and *Drosophila* strongly suggests that *RPD3* is directly involved in

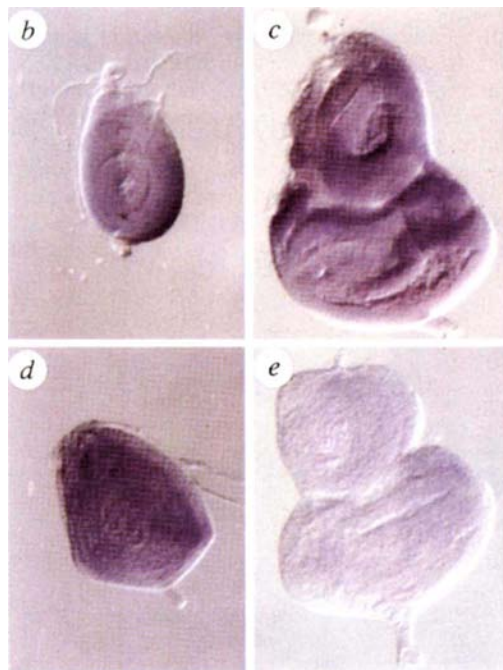
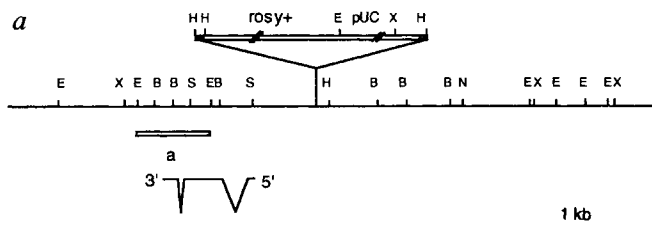


FIG. 2 Transcript accumulation in eye-antenna and leg imaginal discs of mutant and wild-type larvae. a, The 64BC locus. The probe used for *in situ* hybridizations was fragment 'a'. b, Wild-type leg disc; c, wild-type eye antenna disc; d, mutant leg disc; e, mutant eye-antenna disc, in which expression of the gene is strongly reduced.

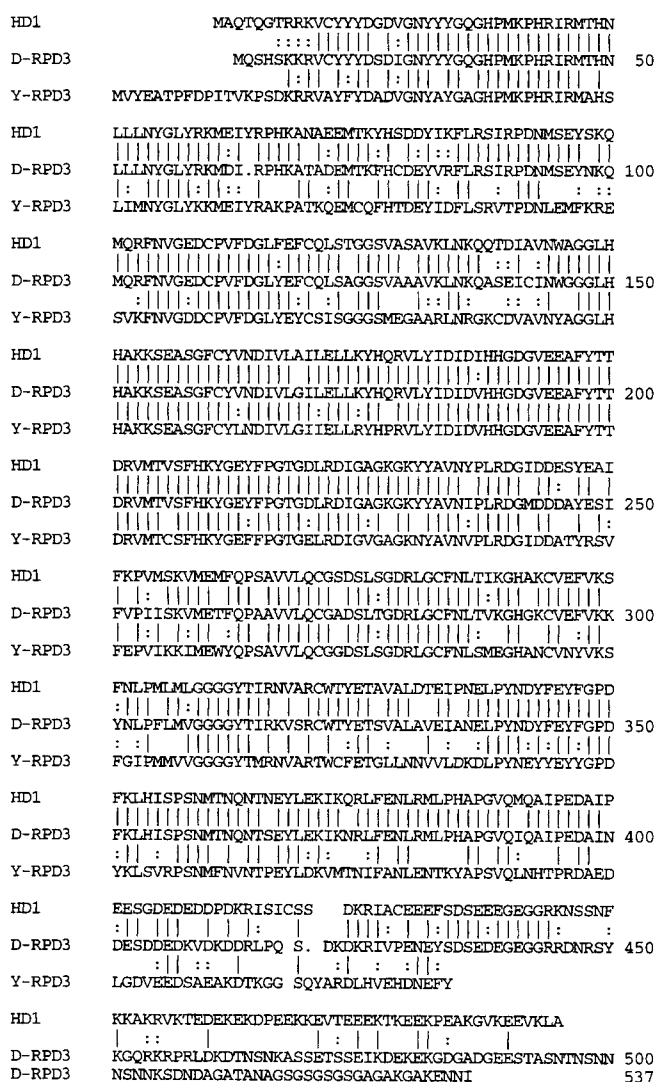


FIG. 3 Comparison of the D-RPD3 deduced protein sequence with the human (HD1) and yeast homologues (Y-RPD3). A vertical bar indicates identical amino acids, and two dots indicate conserved amino acids. In the overlapping region, similarity is 82% (75% identity) with the human homologue, and 71% (58% identity) with yeast. Sequence comparison was done with the BLAST program.

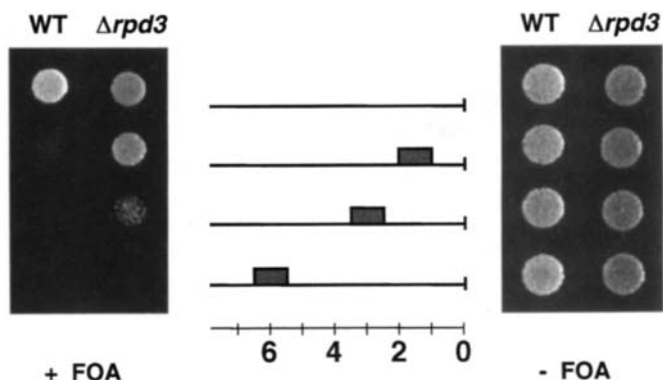


FIG. 4 Effect on yeast telomeric silencing. Parental strain YPH 250 lacks a functional *URA3* gene, whereas strains UCC506, UCC508 and UCC510 contain a *URA3* gene (shaded rectangle). In the three cases the gene is orientated such that transcription proceeds towards the telomere. Cells (10^3) of wild-type (WT) and $\Delta rpd3$ strains were spotted on plates in the presence (left) or the absence (right) of 5-FOA, and were incubated for 2 days at 30 °C.

counteracting genomic silencing.

Yeast and *Drosophila* RPD3 strongly resemble a human histone deacetylase¹¹, and it is likely that the observed enhancement of genomic silencing is due to reduced activity of this enzyme. This conclusion is surprising because histone acetylation (which should be increased when RPD3 function is lost), is generally correlated with increased transcriptional activity^{14–20}. However, there are precedents in yeast and *Drosophila* of such opposite actions, and of putative effects of acetylation on transcriptional repression. In yeast, *rp3* (*sds6*) mutations can restore the metastable repression in strains containing both a *rap1* mutant allele and a mutated silencer element²¹. Moreover, in the silenced mating-type loci, lysine 12 of histone H4 is acetylated, and this acetylation establishes transcriptional silencing²². In this regard, *rp3* mutant strains show increased acetylation of lysines 5 and 12 of histone H4 (ref. 23). In *Drosophila*, a particular isoform of histone H4 acetylated on lysine 12 is abundant in heterochromatin, whereas the other isoforms are underrepresented¹⁸. Lysines 5 and 12 of H4 are acetylated when newly synthesized histone is deposited, and are then rapidly deacetylated²⁴. Here also RPD3, as a hypothetical deposition-related histone deacetylase, could support silencing²².

Thus a specific RPD3-dependent deacetylation could act as a negative signal in establishing a silencing protein complex. More generally, our results suggest that silenced loci may have specialized chromatin structures that are conserved between yeast and *Drosophila*. □

Methods

Whole-mount *in situ* hybridization. This was done using digoxigenin-labelled probes²⁵. The P-element used in the mutagenesis, and plasmid rescue have been described²⁶.

Yeast telomeric silencing. Strains YPH250, UCC506, UCC508 and UCC510 were obtained from D. Gottschling, and have been described⁸. Strains were grown on glucose-minimal media containing casamino acids in the presence and absence of 5-fluoro-orotic acid (1 mg ml⁻¹). To construct Δ *rp3* strains, the RPD3 coding region was obtained by genomic polymerase chain reaction (PCR) and cloned into pUC18. Sequences between the *Nsil* site and *Bam*HI site of RPD3 were then replaced by the *Nsil*-*Bam*HI fragment containing the *HIS3* gene, and the resulting DNA was used to generate Δ *rp3::HIS3* strains by one-step gene displacement.

Received 12 August; accepted 28 October 1996.

- Reuter, G. & Spierer, P. *BioEssays* **14**, 605–612 (1992).
- Karpen, G. H. *Curr. Opin. Genet. Dev.* **4**, 281–291 (1994).
- Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. *Cell* **63**, 751–762 (1990).
- Aparicio, O. M., Billington, B. L. & Gottschling, D. E. *Cell* **66**, 1279–1287 (1991).
- Laurenson, P. & Rine, J. *Microbiol. Rev.* **56**, 543–560 (1992).
- Paro, R. *Curr. Opin. Cell Biol.* **5**, 999–1005 (1993).
- Gartler, S. M. & Riggs, A. D. *Annu. Rev. Genet.* **17**, 155–190 (1983).
- Solter, D. *Annu. Rev. Genet.* **22**, 127–146 (1988).
- Vidal, M. & Gaber, R. F. *Mol. Cell. Biol.* **11**, 6317–6327 (1991).
- McKenzie, E. A. et al. *Mol. Gen. Genet.* **240**, 374–386 (1993).
- Taunton, J., Hassig, C. A. & Schreiber, S. L. *Science* **272**, 408–411 (1996).
- Brown, N. H. & Kafatos, F. C. *J. Mol. Biol.* **203**, 425–437 (1988).
- Renaud, H. et al. *Genes. Dev.* **7**, 1133–1145 (1993).
- Turner, B. M. *J. Cell Sci.* **99**, 13–20 (1990).
- Wolffe, A. P. & Pruss, D. *Cell* **84**, 817–819 (1996).
- Brownell, J. E. & Allis, C. D. *Curr. Opin. Genet. Dev.* **6**, 176–184 (1996).
- Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D. & Broach, J. R. *Genes Dev.* **7**, 592–604 (1993).
- Turner, B. M., Birley, A. J. & Lavender, J. *Cell* **69**, 375–384 (1992).
- Jeppesen, P. & Turner, B. M. *Cell* **74**, 281–289 (1993).
- Brownell, J. E. et al. *Cell* **84**, 843–851 (1996).
- Sussel, L., Vannier, D. & Shore, D. *Genetics* **141**, 873–888 (1995).
- Braunstein, M., Sobel, R. E., Allis, C. D., Turner, B. M. & Broach, J. R. *Mol. Cell. Biol.* **16**, 4349–4356 (1996).
- Rundlett, S. E. et al. *Proc. Natl Acad. Sci. USA* **93**, 13723–13728 (1996).
- Sobel, R. E., Cook, R. G., Perry, C. A., Annunzio, A. T. & Allis, C. D. *Proc. Natl Acad. Sci. USA* **92**, 1237–1241 (1995).
- Masucci, J. D., Miltenberger, R. J. & Hoffmann, F. M. *Genes Dev.* **4**, 2011–2023 (1990).
- Seum, C. et al. *Development* **122**, 1949–1956 (1996).

ACKNOWLEDGEMENTS. We thank A. Spierer for cytogenetic mapping of the locus; C.-H. Tonka for technical assistance; D. Gottschling for yeast strains; and M. Collart for suggesting this collaboration. This work was supported by grants from the Swiss National Science Foundation (P.S. and D.P.), the State of Geneva, and by grants from the National Institute of Health (U.S.) and the German Research Foundation (G.R.).

CORRESPONDENCE and requests for materials should be addressed to P.S. (e-mail: spierer@sc2a.unige.ch). The nucleotide sequence has been deposited in the EMBL Nucleotide Sequence Database, accession no. Y09258.

Crystal structure of the GTPase-activating domain of human p120GAP and implications for the interaction with Ras

Klaus Scheffzek*, Alfred Lautwein†‡, Wolfgang Kabsch†, Mohammad Reza Ahmadian* & Alfred Wittinghofer*

* Max-Planck-Institut für molekulare Physiologie, Rheinlanddamm 201, 44139 Dortmund, Germany

† Max-Planck-Institut für medizinische Forschung, Jahnstr. 29, 69120 Heidelberg, Germany

RAS-RELATED GTP-binding proteins function as molecular switches which cycle between GTP-bound 'on'- and GDP-bound 'off'-states¹. GTP hydrolysis is the common timing mechanism that mediates the return from the 'on' to the 'off'-state. It is usually slow but can be accelerated by orders of magnitude upon interaction with GTPase-activating proteins (GAPs). In the case of Ras, a major regulator of cellular growth, point mutations are found in approximately 30% of human tumours which render the protein unable to hydrolyse GTP, even in the presence of Ras-GAPs. The first structure determination of a GTPase-activating protein reveals the catalytically active fragment of the Ras-specific p120GAP (ref. 2), GAP-334, as an elongated, exclusively helical protein which appears to represent a novel protein fold. The molecule consists of two domains, one of which contains all the residues conserved among different GAPs for Ras. From the location of conserved residues around a shallow groove in the central domain we can identify the site of interaction with Ras-GTP. This leads to a model for the interaction between Ras and GAP that satisfies numerous biochemical and genetic data on this important regulatory process.

Five mammalian Ras-specific GAPs are known to date (summarized in ref. 3), in addition to several others from different organisms. Although of various sizes and domain composition they all contain the catalytic module, also called GAP-related domain, with characteristic blocks of sequence homology (Fig. 1a). p120GAP, the first GAP to be discovered², is a protein of relative molecular mass 120,000 (*M_r* 120K), consisting of several signalling domains^{4,5} (Fig. 1b). Disruption of the gene in mice is embryonically lethal owing to vascular-system defects and neuronal apoptosis⁶. Neurofibromin (NF1), another Ras-GAP, is the product of the presumed tumour-suppressor gene *NF1* that is mutated or disrupted in patients suffering from type 1 neurofibromatosis⁷. Somatic mutations of the *NF1* gene are also found in solid tumours not associated with neurofibromatosis⁸. The importance of the proper functioning of the GTPase cycle is thus demonstrated by the occurrence of human diseases associated with mutations in either Ras itself or the regulatory proteins, both of which affect the rate of GTP hydrolysis and thus the regulation of cell growth. As a first step to investigate how GAPs might interact with Ras and stimulate Ras-mediated GTP hydrolysis, we have determined the crystal structure of GAP-334, a catalytically active fragment containing the 334 C-terminal residues of human p120GAP.

The structure of GAP-334 was determined with the multiple isomorphous replacement (MIR) method using mercury, gold and lead derivatives of crosslinked crystals⁹ to solve the phase problem at 3 Å resolution. By using synchrotron radiation combined with cryocrystallography we obtained X-ray data of native crystals,

‡ Present address: Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK

which allowed structure refinement at 1.6 Å resolution with a current *R*-factor of 26.7% for a model comprising 324 residues of the GAP-334 sequence and 95 water molecules (Table 1). Refinement is still in progress; as the *R*-factor for a data set of 2.4 Å resolution is 22.5%, we do not expect significant changes in the final model. A representative segment of electron density is shown in Fig. 2a.

GAP-334 is an elongated curved molecule of approximate size 70 × 30 × 30 Å, and contains exclusively helical secondary structure elements (Fig. 2b, c). It seems to be composed of two domains: a central domain, GAP_c, which makes up two-thirds of GAP-334 and presumably represents the catalytic domain, and an extra domain, GAP_{ex}. Residues 714–763 from the N-terminal portion and 982–1047 from the C-terminal portion of the protein form three α-helices each, which come together to form GAP_{ex}. The arrangement of the helices is stabilized mainly by interactions between the side chains of hydrophobic amino acids. GAP_c consists of 218 residues located between Lys 764 and Pro 981 from the middle portion of the GAP-334 sequence. It contains eight α-helices, arranged such that the approximately antiparallel helices α6_c and α7_c, which are connected by a large loop, form the bottom of a shallow groove on the surface of the protein. Because these two helices contain most of the invariant residues of GAP, and for reasons outlined below, we consider this to be the active site of the protein, and refer to it as the 'catalytic track' because of its appearance in the model. The presumed active-site groove is bordered on three sides by helices and loops which stabilize the structure, keeping the catalytic track open to the solvent (Fig. 2b, c).

GAP_c corresponds to what has been found by proteolysis experiments to be a minimum catalytic domain of NF1 that retains full GAP activity³. In contrast, the minimum soluble catalytic fragment of GAP-334 that can be obtained by limited proteolysis

or molecular cloning contains 46 additional N-terminal residues (Fig. 1a). GAP_{ex} and GAP_c contact each other by means of a hydrophobic interface formed by residues from α2_{ex}, α3_{ex}, α1_c and α2_c. Although the function of the extra domain is not obvious, its removal would expose hydrophobic residues to the solvent, perhaps rendering the protein insoluble, as was observed for p120GAP (ref. 3). The structure supports the observation that the large loop connecting the extra and central domains in the C-terminal part of the polypeptide chain is cleaved by proteases in both NF1 and p120GAP (ref. 3). Of 14 prolines in GAP-334, 12 are located within GAP_c, two of which are invariant and one highly conserved. Except for Pro 912 in block 3A, we find them in loops connecting the helices, usually as additional contributors to the structural integrity of the helical pattern. Because the N- and C-terminal ends of GAP_c are situated close to each other, and the invariant and highly conserved residues of all Ras-GAPs are located inside this domain, we assume that this structural module is found in other, if not all, Ras-GAPs.

It had been found that an in-built GAP domain in the α subunits of heterotrimeric G proteins is the striking difference between these and the Ras-like proteins¹. As well as the G-domain common to both, an α-helical insertion is present in Gα proteins which has been shown to provide GAP activity^{10–12}. The observations that Gα proteins complexed with GDP bind the complex ion AlF₄[−], thereby forming a transition-state analogue, but that Ras can do so only in the presence of its GAP¹³, suggested that the structure of GAP-334 might resemble that of the helical domain in Gα proteins. However, we failed to detect significant similarity between these structures.

GAP_c carries all of the invariant and highly conserved residues (except Leu 757) of Ras-GAPs, most of which are found in the three blocks of sequence homology (Figs 1a, 3a). Block 1 contains the invariant Arg 789, located in a short 3₁₀-helix as part of the

TABLE 1 Summary of crystallographic analysis

Crystal: spacegroup $P2_12_12_1$; $a = 42.2 \text{ Å}$, $b = 55.6 \text{ Å}$, $c = 142.2 \text{ Å}$, $\alpha = \beta = \gamma = 90^\circ$.						
	Native 1*	Native 2*	KAu(CN) ₂	K ₂ Hgl ₄	EMP†	TMLA†
			(soak, 4 mM, 42 h)	(cocryst., 3:1)	(soak, 0.1 mM, 12 h)	(soak, 1 mM, 48 h)
Data collection						
Resolution (Å)	1.6	2.9	3.5	3.0	3.5	3.0
Observed reflections	123791	25121	11876	14326	9041	10944
Unique reflections	41352	7298	4483	5620	4156	6053
Completeness (%)	96.1	92.4	98.5	78.7	87.6	84.2
R_{sym} (%)‡	7.7	5.1	8.3	9.2	8.1	4.1
R_F (%)§			22.4	31.2	21.1	13
MIR analysis						
Resolution (Å)		15–3	15–3.5	15–3.4	15–4	15–3
Sites			2	1	2	1
$R_c $			0.62	0.67	0.84	0.7
$F_w/E $			0.95	0.85	0.45	0.77
Mean f.o.m.		0.33				
Refinement						
Resolution (Å)	6–1.6					
R_{cryst} (%)#	26.7					
R_{free} (%)☆	34					
R.m.s. bond length (Å)	0.012					
R.m.s. bond angle (deg)	1.5					

* Data sets of native crystals.

† EMP, ethyl mercury phosphate; TMLA, tri methyl lead acetate.

‡ $R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \bar{I}_h| / \sum_h \bar{I}_h$, where I_{hi} is the scaled intensity of the i th symmetry related observation of reflection h , and $\bar{I}_h$ is the mean value.

§ $R_F = 2 \sum_h |F_{\text{PH}} - F_P| / \sum_h |F_{\text{PH}} + F_P|$, where F_{PH} and F_P are the derivative and native structure amplitudes.

|| $R_c = \sum_h |F_{\text{PH,obs}} - F_{\text{PH,calc}}| / \sum_h |F_{\text{PH,obs}} + F_{\text{PH,calc}}|$.

¶ Phasing power as defined by $F_w/E = (f_h^2 / \sum_h (F_{\text{PH,obs}} - F_{\text{PH,calc}})^2)^{1/2}$; f_h is the heavy-atom scattering factor.

$R_{\text{cryst}} = \sum_h |F_{\text{oh}} - F_{\text{ch}}| / \sum_h F_{\text{oh}}$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes for reflection h , respectively.

☆ Value of R_{cryst} calculated for 10% randomly chosen reflections not included in the refinement.

loop between helices α_{1c} and α_{2c} . Mutations of the corresponding residue in neurofibromin and p120GAP interfere with GAP activity without having a major effect on Ras binding¹⁴ (M.R.A., unpublished data). The highly conserved Glu 777 interacts with the hydroxyl group of Thr 792, which is conserved as either threonine or serine in other Ras-GAPs. This interaction might be important for protein stabilization, thus explaining the conservation of both residues, a conclusion supported by mutagenesis studies¹⁴. The corresponding neurofibromin mutant E1264Y has reduced GAP activity *in vitro* but is functional in yeast, supporting the conclusion that it is not involved in catalysis¹⁵. Glu 829 between block 1 and 2, and Arg 928 in block 3B, both conserved as negatively or positively charged residues, respectively, are forming a salt bridge that contributes to protein stability¹⁴.

The block 2 region, which contains conserved prolines (Pro 867 and Pro 886) at either end of α_{5c} , seems to be important for structural integrity rather than being directly involved in the

interaction between Ras and GAP. Mutations affecting the activity have been found in this part but no studies implicating this region to be directly involved in catalysis have been reported. Incidentally, in neurofibromin an insertion of 21 amino acids following Gln 1373 in this region (Fig. 1a) has been identified as the product of alternative splicing¹⁶, and it was shown that the protein encoded by the type II transcript has slightly reduced GAP activity¹⁷. A corresponding insertion in GAP-334 would be located in helix α_{5c} and, as judged from the structure, would not necessarily interfere with GAP action.

Residues from the most highly conserved block 3 form the catalytic track, helices α_{6c} and α_{7c} with a long connecting loop (Figs 2b, c and 3a). They contain most of the invariant amino acids, including the fingerprint ⁹⁰¹FLR...PA...⁹¹² motif diagnostic for Ras-GAPs. The FLR consensus motif (residues 901–903), which is preceded and followed by other highly conserved hydrophobic amino acids, is conspicuously located where helix α_{6c} is

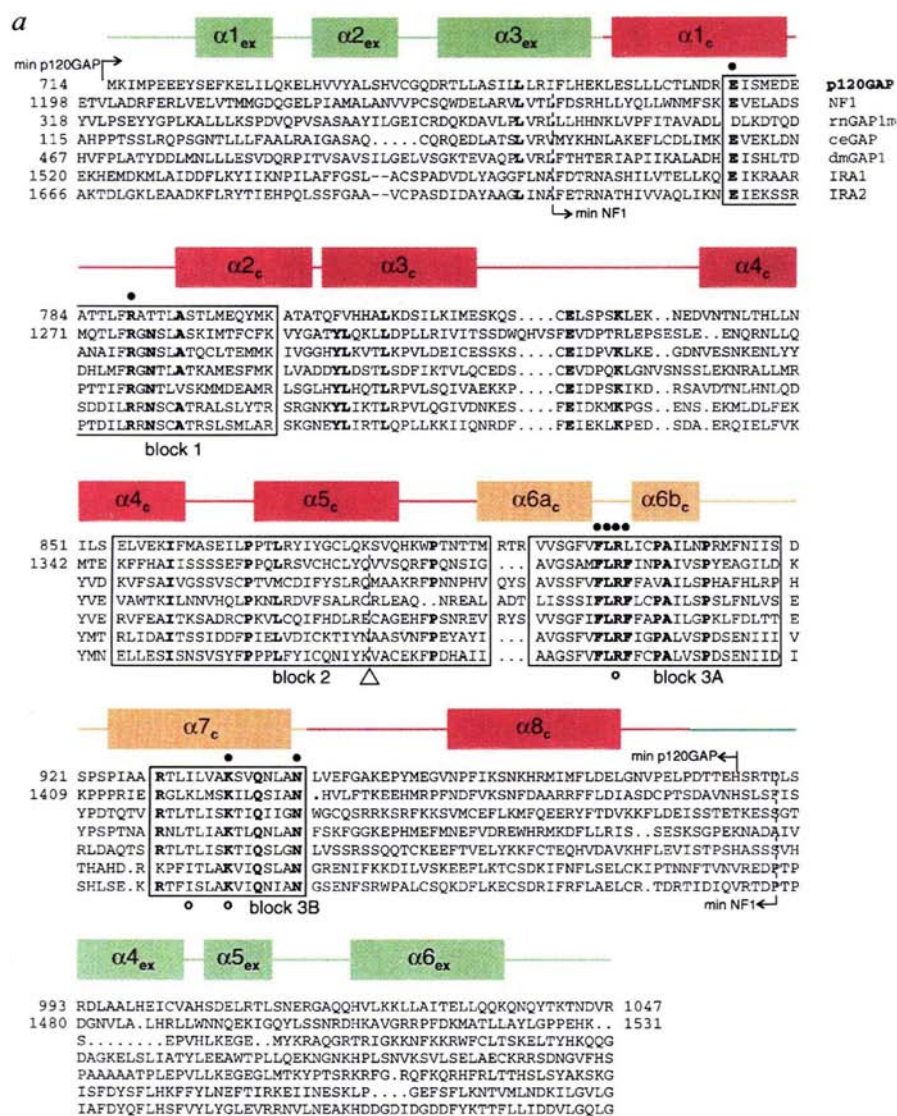


FIG. 1 Architecture of GAP-334 in terms of sequence and secondary structure. **a**, Secondary structure assignment (program DSSP; ref. 34) of GAP-334 (GenBank accession number M23379), and sequence alignment (program Pileup, Pretty GCG, Madison, WI; adjusted manually for maximum homology) of Ras-GAPs from different organisms. NF1 is neurofibromin, the protein product of the neurofibromatosis type I gene (M89914); mGAP1m (D30734) is the rat homologue of the drosophila GAP gene dmGAP1 (M86655); ceGAP is the *Caenorhabditis elegans* GAP gene (Z46266); and IRA1 (M24378) and IRA2 (M33779) are the yeast *Saccharomyces cerevisiae* homologues of GAP. Minimal domains of p120GAP and neurofibromin that can be expressed stably in *Escherichia coli* and retain full GTPase-activating activity are marked as 'min NF1' and 'min p120GAP', respectively, as identified by proteolysis experiments³. Open triangle, the location of the alternatively spliced transcript in neurofibromin; filled circles, residues mutated in p120GAP or NF1 as discussed in the text; open circles, residues mutated in NF1 patients or solid tumours. **b**, Domain composition of p120GAP, showing the C-terminal location of the GAP-related domain (GRD) in relation to the SH2, SH3, PH and CalB domains.

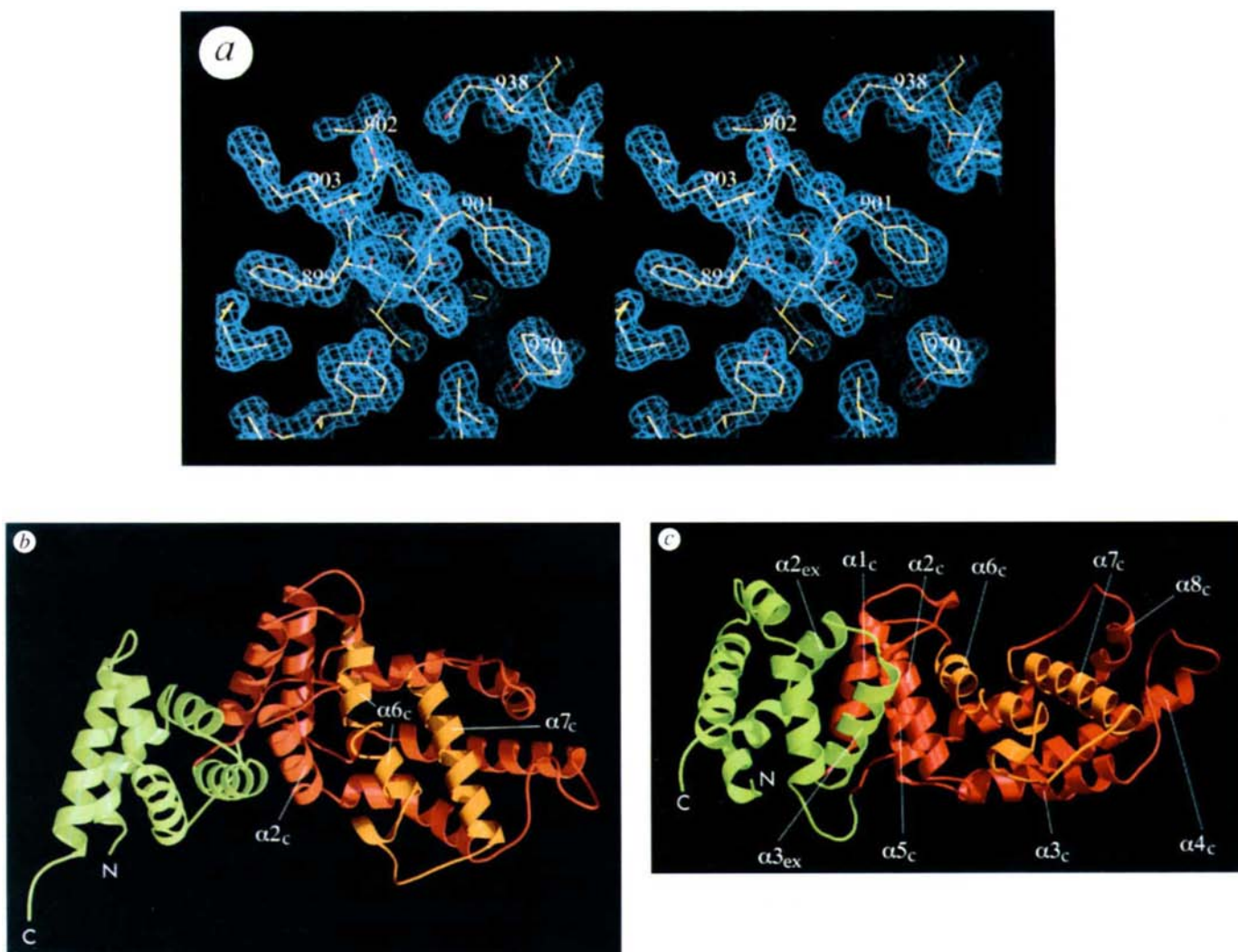


FIG. 2 The three-dimensional structure of GAP-334. *a*, $2F_o - F_c$ electron-density map contoured at 13% of the maximum showing the region around the four-residue turn $^{901}\text{FLRL}^{904}$ inserted in $\alpha 6_c$ which is part of the Ras-GAP fingerprint motif FLR...PA..P, viewed down the helical axis. *b*, Ribbon representation (drawn with MOLSCRIPT; ref. 35) of GAP-334, with the

extra domain GAP_{ex} in green, the central-catalytic domain GAP_c in red, and the catalytic track ($\alpha 6_c$ and $\alpha 7_c$ in block 3) highlighted in orange. *c*, As *b*, but with the molecule rotated by approximately 90° about a horizontal axis in the plane of the paper, revealing the active-site groove in the surface of GAP_c .

interrupted by the insertion of one residue, such that it apparently contains a four-residue turn. The hydrophobic residues (Phe 899, Val 900, Leu 904 and Ile 905) close to the FLR motif, including Phe 901 itself, contribute to different hydrophobic cores on either side of the catalytically important Leu 902 and Arg 903 (Fig. 3*a-c*). Mutations of Phe 901 from FLR to PA..P, a phenylalanine in all other Ras-GAPs) destabilize the protein, supporting its role in core stabilization^{14,18}.

In contrast to the hydrophobic residues mentioned, Leu 902 has been demonstrated to be important for the functional interaction between GAP and Ras. The mutation $\text{L} \rightarrow \text{I}$ at position 902 drastically reduces the catalytic activity of GAP by affecting k_{cat} , without grossly disturbing the affinity for Ras¹⁹. As the protein could be expressed stably in *Escherichia coli*, it was concluded that Leu 902 is not involved in core formation but rather in the interaction between GAP and Ras. This is also supported by the structure, as Leu 902, as part of the four-residue helical turn, is exposed to solvent, its side chain pointing into the interior of the shallow groove. Arg 903, part of the FLR motif, is held in place by several interactions, most importantly between its guanidinium group and the main-chain carbonyl oxygens of Phe

788, Arg 789 and Ala 790. Arg 903 is important for catalysis, and its replacement by glutamate or lysine reduces catalytic efficiency with only a minor effect on binding¹⁹. The mutation of the corresponding arginine to methionine in neurofibromin (R1391M) has only a minor effect on the binding of Ras, but renders the protein unable to form a transition-state analogue of the GTPase reaction¹³.

Helix $\alpha 7_c$, the second helix of the catalytic track, with three invariant residues (Lys 935, Gln 938 and Asn 942) at its C-terminal end, has been implicated in Ras binding^{14,20} (M.R.A., unpublished data). Because these invariant residues in $\alpha 7_c$ are situated 3–4 residues apart, corresponding to one helical turn, they are lined up along one face of this helix and point towards the postulated active-site groove. Lys 1423 in NF1 (corresponding to Lys 935 of p120GAP) has been found mutated to glutamate or glutamine in solid tumours and in patients suffering from type I neurofibromatosis^{7,8}; it has been claimed that these mutations affect the enzymatic reaction rather than the binding of Ras. Other reports have shown that various mutations of Lys 1423 affect the binding of Ras and/or the stability of the protein^{14,21,22}. The three-dimensional structure of GAP provides an explanation for the high conserva-

tion of Ala 941; its methyl group interacts with the phenylalanine of the FLR motif, and larger side chains could not be accommodated in the hydrophobic environment of the region between α_6 and α_7 .

Many mutational studies have suggested regions in Ras and GAP that participate in the interactions necessary for the GAP-accelerated GTPase reaction. The sites on Ras have been mapped extensively, and because GAP recognizes the GTP-bound form of Ras specifically, it is not surprising that it involves residues of the so-called switch regions switch I and switch II, which have been shown to be conformationally different in the GTP-bound and the GDP-bound states. A third region of contact between Ras and GAP is the region 89–102 (refs 23, 24), with Asp 92 being a crucial residue at this site of interaction²⁰. Naturally occurring mutations of Gly 12, Gly 13 and Gln 61 activate the oncogenic potential of Ras, owing to a failure of GAP to activate their GTPase activity, suggesting that residues in positions 12, 13 and 61 also affect the interaction between Ras and GAP. Because these mutants bind to GAP with roughly normal affinity, these substitutions would seem to affect other steps of the reaction pathway¹³. Even small structural perturbations of the P-loop in Ras (residues 10–17) are not tolerated by GAP, suggesting that there is a close contact between Gly 12, Gly 13 and GAP, probably in the transition state of the reaction¹³.

Mutagenesis studies on neurofibromin^{7,13,15,21,22} and p120 GAP^{14,18,19} and peptide-binding studies²⁵ have led to the mapping

of residues that are critical for a successful GAP–Ras interaction to the block 1 and 3 region. We have used this mapping in our attempts to dock Ras·GTP²⁶ onto the surface of GAP-334 (Fig. 4). It is remarkable how the curvature of the groove in GAP-334 can be filled by the Ras model in the particular orientation shown; the electrostatic surfaces of Ras and GAP-334 (not shown) are also complementary for this docking model. More importantly, all the requirements for this interaction seem to be satisfied by the suggested complex formation. Four regions of Ras would be involved in this presumed protein–protein interaction; the P-loop containing Gly 12; switch I containing the effector residues; switch II containing Gln 61 and the regions around Asp 92 on Ras; and the block 1 and 3 regions on GAP-334. Residues located after α_7 , including Lys 948, are aligned with residues of the effector region in Ras. Residues in the corresponding region of neurofibromin have been found as suppressors of a Lys 1423-inactivating mutation²¹ or of morphological transformation by oncogenic Ras, owing to tighter binding²⁷, which emphasizes the plausibility of the docking model. Thus, although atomic details of the presumed interaction between Ras and GAP cannot be extracted from our model, it should be useful for suggesting further experiments to study this interaction. Whatever the exact atomic details, it is already obvious that Ras uses two different ways of binding to its regulators and effectors: in the complex with the effector Raf, an inter-protein β -sheet is formed²⁸, whereas the interaction with the helical protein GAP probably uses a multicontact surface involving β -strands and helices.

To understand the mechanism of the Ras–GTPase reaction, we need to know how this reaction is stimulated 10^5 -fold by GAP. Because the catalytic reaction is taking place on Ras itself, the question has centred on two possible models for the acceleration. In one model the machinery for efficient hydrolysis is entirely on Ras, and the role of GAP is to promote the formation of a GTPase-competent state, and experiments suggesting or questioning such a role have been reported^{13,29,30}. In the second model (the 'arginine-finger hypothesis'), GAP supplies residues into the active site of Ras where they participate in catalysis, probably by stabilizing the transition state. Such residues could be arginines, which in G α proteins^{10,11} and metabolic enzymes such as

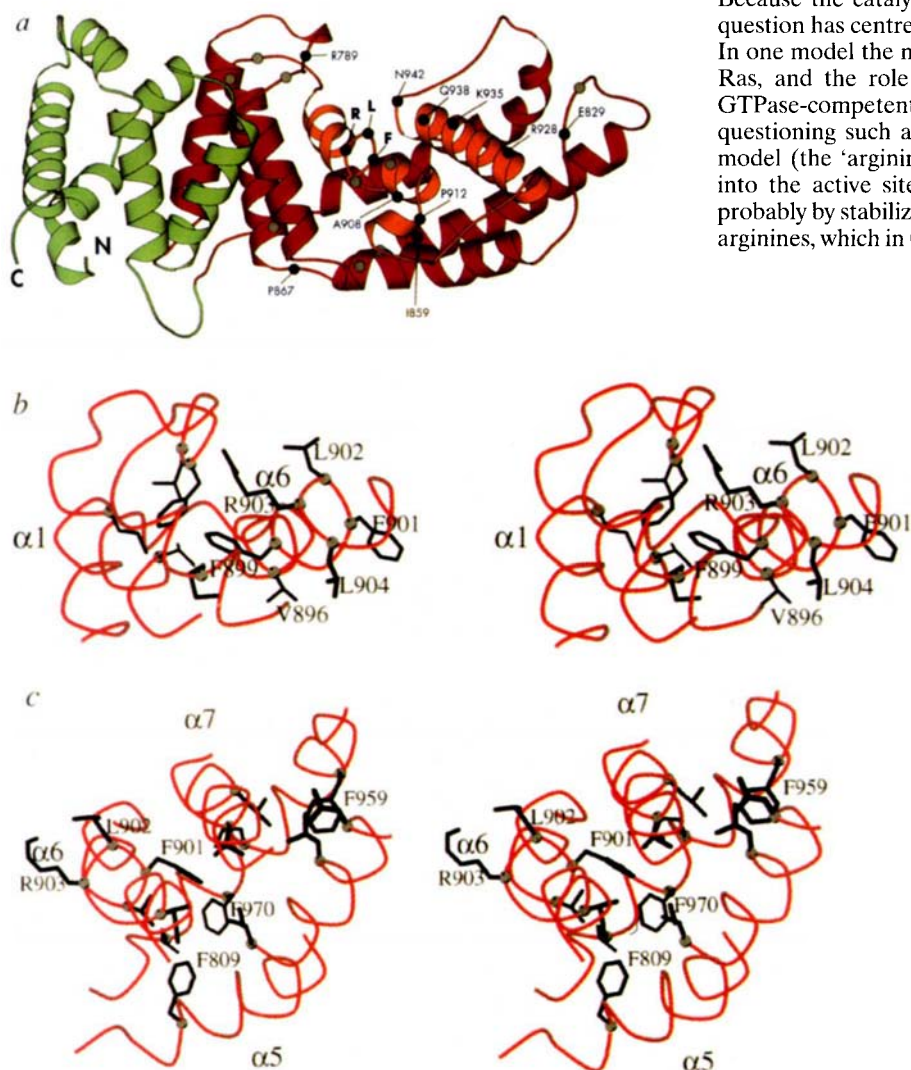


FIG. 3 Details of the structure. *a*, Ribbon plot with a similar orientation to Fig. 2c, showing the position of the invariant and highly conserved residues as black and grey dots, respectively. *b*, *c*, Stereo views demonstrating how residues on the left (in *b*) or right (in *c*) side of the totally invariant fingerprint 901FLR903 region (with respect to the orientation in *a*) participate in the formation of two different hydrophobic cores. This arrangement presumably stabilizes the position of the apparent four-residue turn in α_6 and positions Leu 902 and Arg 903 for catalysis.

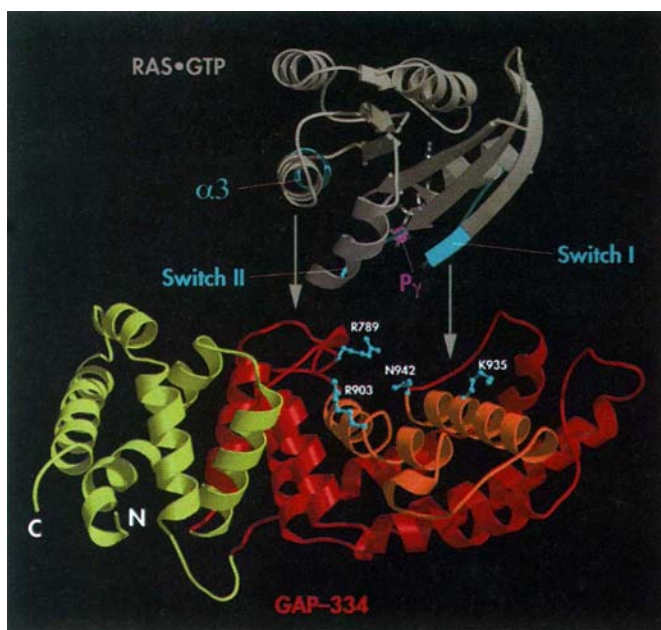


FIG. 4 Docking approach between Ras and GAP, modelled manually on the basis of numerous genetic and biochemical studies on Ras and GAPs. Ras (grey) is approaching GAP such that its regions (blue) implicated to be involved in GAP-accelerated GTP hydrolysis contact the catalytic domain of GAP-334, and that the γ -phosphate (P_{γ} , purple) of GTP is close to residues suggested to participate in the enzymatic reaction.

adenylate kinases have been shown to have such a function. The 'arginine-finger' hypothesis has been supported by the recent finding that Ras needs stoichiometric amounts of GAP with at least one of the invariant arginines¹³, for it to bind AlF_4^- and form a presumed transition-state analogue.

The structure of the catalytic domain of GAP shows that both arginines are in proximity both to each other and to the γ -phosphate in the (albeit imprecise) docking model of the Ras-GAP complex (Fig. 4). It is thus conceivable that they both contribute to catalysis without the requirement for extensive intramolecular rearrangements. The guanidinium group of Arg 903 contacts the main chain in the region containing Arg 789, raising the possibility that one arginine (Arg 789) is used for catalysis while the other (Arg 903) stabilizes its orientation. A comprehensive understanding of how GAP accelerates Ras-mediated GTP hydrolysis, why it is unable to do so in oncogenic mutants, whether it will be possible to restore the GTPase activity of such mutants, and how other domains of p120GAP modulate catalytic efficiency³, awaits further structural studies of the communication between Ras and GAP. An important step towards this goal has been made by our determination of the X-ray structure of the catalytic fragment of p120GAP. □

Methods

Data collection. Expression and purification of recombinant GAP-334, crystal growth from hanging drops and improvement of the crystals by microseeding have been described⁹. Diffraction data were recorded by the rotation method (crystal detector distance, 13 cm; oscillation range, 0.0833°; X-ray source, rotating anode Elliott/GX18, 35 kV/50 mA, CuK_{α} radiation; detector, Siemens/Nicolet area detector) and processed with XDS³¹. Native high-resolution data were collected using synchrotron radiation at the EMBL outstation (DESY, BW7B, Hamburg), at ESRF (BL 4, Genoble), and at NSLS (X12C, Brookhaven). The best data set of 1.6 Å resolution (Native 1) was obtained from a single crystal, flash frozen in liquid nitrogen using 30% PEG 3350 as cryo protectant,

exposed in the high-brilliance beam of BL4 at ESRF (crystal detector distance, 20 cm; oscillation range, 0.7°; detector, MAR image plate).

Multiple isomorphous replacement. Heavy-atom derivatives were prepared by soaking crosslinked crystals⁹ in 400 μ l of reservoir solution containing 0.05–5 mM of the respective compound or by co-crystallization with subsequent crosslinking. Of 25 compounds tested, we obtained 4 derivatives of quality sufficient for phase determination, using a native data set of a crosslinked crystal (Native 2) as reference in MIR analysis. Derivative data were analysed with a program suite by W. Kabsch (unpublished), resulting in a heavy-atom model consisting of three distinct sites of various occupancies. An electron-density map was calculated at 3 Å resolution and inspected visually after solvent flattening. This map revealed GAP-334 as an elongated helical molecule, and was consistent with the assumption of one protein molecule in the asymmetric unit.

Model building and refinement. For the convenience of model building (using the program O (ref. 32)), poly-alanine helices were positioned manually in the map and served as a frame for subsequent chain tracing within electron density of sufficient connectivity. Refinement (using XPLOR version 3.1 (ref. 33)) of a poly-alanine model consisting of 12 segments and 227 residues (data between 10 and 3 Å) gave an R-factor of 44%. The resulting model fitted well into the initial map, and after thorough inspection we found a sequence register along connective electron density which was consistent with heavy-atom positions being close to cysteine residues of the protein. Starting from the respective positions, alanines were replaced by the residues of the GAP-334 sequence where they fitted into electron density. In the following rounds of alternate model building and refinement (all done with O and XPLOR version 3.1), during which resolution was gradually increased, the model was extended to 324 residues corresponding to residues 718–1041 of p120GAP. The 4 N-terminal and 6 C-terminal residues are not visible in the electron density; regions 834–838 and 982–989 presently show poor electron density and were modelled stereochemically. The model contains 95 water molecules; refinement of the structure is still in progress.

Received 10 September; accepted 19 November 1996.

- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349**, 117–127 (1991).
- Trahey, M. & McCormick, F. *Science* **238**, 542–545 (1987).
- Ahmadian, M. R., Wiesmüller, L., Lautwein, A., Bischoff, F. R. & Wittinghofer, A. *J. Biol. Chem.* **271**, 16409–16415 (1996).
- Vogel, U. et al. *Nature* **335**, 90–93 (1988).
- Trahey, M. et al. *Science* **242**, 1697–1700 (1988).
- Henkemeyer, M. et al. *Nature* **377**, 695–701 (1995).
- Xu, G. et al. *Cell* **62**, 599–608 (1990).
- Li, Y. et al. *Cell* **69**, 275–281 (1992).
- Scheffzek, K., Lautwein, A., Scherer, A., Franken, S. & Wittinghofer, A. *Proteins Struct. Funct. Genet.* (in the press).
- Coleman, D. E. et al. *Science* **265**, 1405–1412 (1994).
- Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. *Nature* **372**, 276–279 (1994).
- Markby, D. W., Onrust, R. & Bourne, H. R. *Science* **262**, 1895–1901 (1993).
- Mittal, R., Ahmadian, M. R., Goody, R. S. & Wittinghofer, A. *Science* **273**, 115–117 (1996).
- Miao, W., Eichelberger, L., Baker, L. & Marshall, M. S. *J. Biol. Chem.* **271**, 15322–15329 (1996).
- Gutmann, D. H. et al. *Oncogene* **8**, 761–769 (1993).
- Nishi, T. et al. *Oncogene* **6**, 1555–1559 (1991).
- Rey, I., Taylor-Harris, P., van Erp, H. & Hall, A. *Oncogene* **9**, 685–692 (1994).
- Hettich, L. & Marshall, M. *Cancer Res.* **54**, 5438–5444 (1994).
- Brownbridge, G. G., Lowe, P. N., Moore, K. J. M., Skinner, R. H. & Webb, M. R. *J. Biol. Chem.* **268**, 10914–10919 (1993).
- Morcos, P., Thapar, N., Tusneem, N., Stacey, D. & Tamanoi, F. *Mol. Cell. Biol.* **16**, 2496–2503 (1996).
- Poulet, P., Lin, B., Esson, K. & Tamanoi, F. *Mol. Cell. Biol.* **14**, 815–821 (1994).
- Wiesmüller, L. & Wittinghofer, A. *J. Biol. Chem.* **267**, 10207–10210 (1992).
- Parni, M. C., Bernardi, A. & Parmeggiani, A. *EMBO J.* **15**, 1107–1111 (1996).
- Yoder-Hill, J., Golubic, M. & Stacey, D. W. *J. Biol. Chem.* **270**, 27615–27621 (1995).
- Rubinfeld, B. et al. *Int. J. Pept. Protein Res.* **38**, 47–53 (1991).
- Pai, E. F. et al. *EMBO J.* **9**, 2351–2359 (1990).
- Mori, S. et al. *J. Biol. Chem.* **270**, 28834–28838 (1995).
- Nassar, N. et al. *Nature* **375**, 554–560 (1995).
- Moore, K. J. M., Webb, M. R. & Eccleston, J. F. *Biochemistry* **32**, 7451–7459 (1993).
- Rensland, H., Lautwein, A., Wittinghofer, A. & Goody, R. S. *Biochemistry* **30**, 11181–11185 (1991).
- Kabsch, W. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, T. A. *XPLOR version 3.1* (Yale University, 1992).
- Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
- Kraulis, P. J. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

ACKNOWLEDGEMENTS. We thank I. Schlichting for help with synchrotron data collection and cryo-crystallography; A. Becker for advice on heavy-atom chemistry; the staff at the synchrotron facilities of ESRF, EMBL and NSLS for support; L. Wiesmüller, T. Schweins and M. Geyer for discussion; F. Schmitz for help with graphical representation; R. Mittal for critical reading of the manuscript; R. Scheubach for secretarial help; and K. Holmes for support. K.S. acknowledges financial support from the Peter and Traudl Engelhorn Stiftung (Germany) and the National Neurofibromatosis Foundation (USA).

CORRESPONDENCE and requests for materials should be addressed to A.W. (e-mail: alfred.wittinghofer@mpi-dortmund.mpg.de). Coordinates have been deposited in the Brookhaven Protein Databank, accession number 1WER, where they will be held for one year.

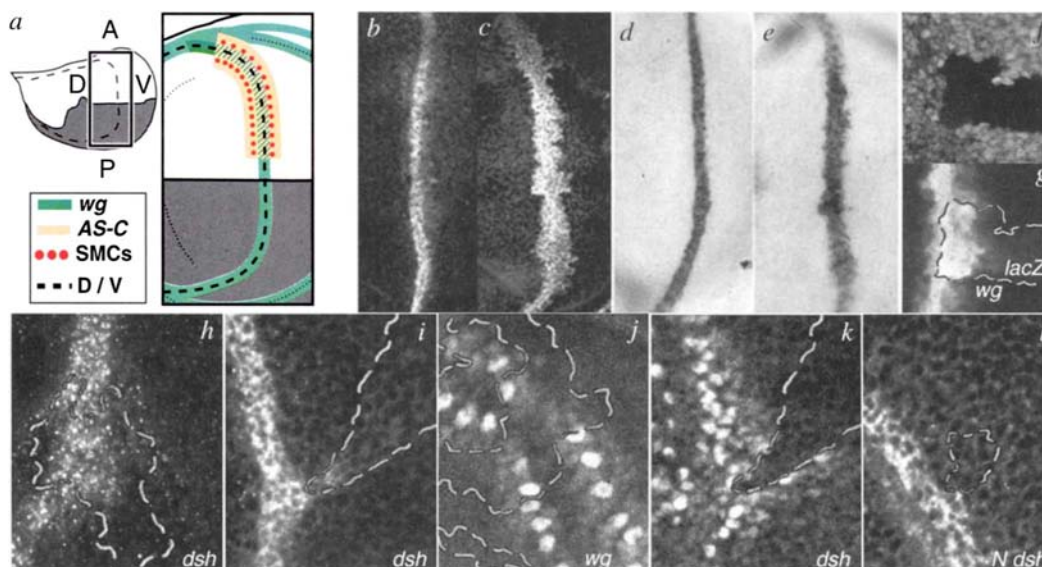
wingless refines its own expression domain on the *Drosophila* wing margin

Eric J. Rulifson, Craig A. Micchelli, Jeffrey D. Axelrod, Norbert Perrimon & Seth S. Blair

Nature 384, 72–74 (1996)

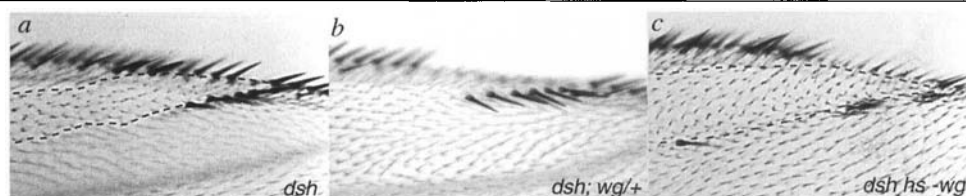
BECAUSE of an error in the production process all the figures in this Letter were unclear and of poor quality. They are reproduced again below.

FIG. 1 a–i, Margin *wg* expression after loss of *wg* or *dsh* functions. a, Diagram of late third instar wing disc. Box outlines margin region shown in the remaining figures, with axes and patterns of gene expression as marked (see text for details). AS-C, *acheate-scute* complex expression; DV, dorso-ventral. b–e, *wg*^{ts} (*wg*^{ts}/*wg*^{ts}) margins, stained for Wg protein (b, c) or messenger RNA (d, e) expression. b, d, At permissive temperature, protein and mRNA expression was normal. c, e, After a 12 h shift to restrictive temperature, protein expression expanded to a region twice the normal width (protein: 55/57 discs; mRNA: 17/19 discs). f, g, Hypomorphic *wg*^{ac2} clone on the margin, shown by the absence of anti-Myc staining (f; in g and subsequent panels clones are marked by a dotted outline and *). Clones caused expansion of *wg*-*LacZ* expression (anti-β-gal; 72/97, none showed loss). h, i, *dsh* clones that intersected (h) or sat immediately adjacent to (i) the *wg* stripe elevated anti-Wg-staining cell-autonomously (*dsh*⁷⁵: 15/17; *dsh*²⁶: 12/12). Similar effects were observed in anterior, posterior, dorsal and ventral clones. In some cases, it seemed that the normal *wg* stripe expanded or distorted to meet the ectopic anti-Wg staining within clones away from the margin (see i; 7 clones). It is unclear whether ectopic *wg* expression is being induced in wild-type cells between the clone and the margin, or if cells at the margin are distorting or rearranging near the clone. j, k, Anti-Scute or Achaete staining in *wg*^{ts} and *dsh*^{ts} clones. j, *wg*^{ts} clones that intersected the *wg* stripe showed non-autonomous loss of Sc (18/22 large clones), as did 5/7 *wg*^{ac2} clones (not shown). Sc was expressed at wild-type levels in *wg* mutant cells that were 1–2 cell diameters from the clone boundary, presumably in response to Wg secreted by wild-type cells. Most small clones, where all cells were



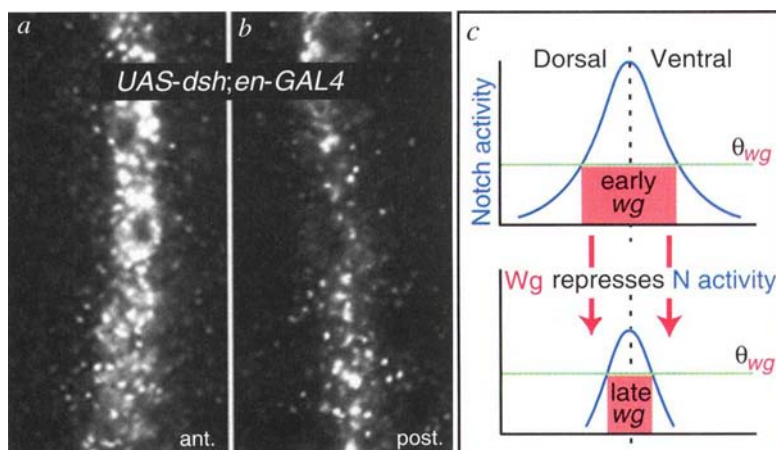
close to the boundary, showed normal Sc levels. k, *dsh*^{ts} clones within the margin proneural region showed cell-autonomous loss of sc expression (*dsh*⁷⁵: 9/9; *dsh*²⁶: 15/15). *dsh* clones that intersected the *wg* stripe, and produced ectopic *wg*, also generated ectopic sc expression outside the normal proneural region in *dsh*^{ts} cells near the clone boundary (arrow; note *wg* expression in same clone in i). l, Margin *wg* expression after loss of *Notch* (*N*) and *dsh*. *Notch*^{ts} *dsh*^{ts} clones that lay adjacent to (l) or intersected the *wg* stripe (not shown) showed cell-autonomous loss of anti-Wg staining without expansion of *wg* expression (29/29 clones). *Notch*^{ts} *dsh*^{ts} clones in the anterior also lose anti-Sc staining (E.J.R., C.A.M., M. Halevy and S.S.B., manuscript in preparation).

FIG. 2 *dsh*^{ts} clones induced ectopic bristles off the wing margin in a *wg*-dosage-dependent fashion. a, *dsh*^{ts} clone in adult wing, marked with *y* and *F^{66a}* (dotted outline), extending from the margin into the interior of the wing blade. The average distance from farthest bristle to margin was 4.5 cells (± 1.0 , range 3–7, *n* = 23). b, *dsh* clone induced in a *wg*^{G22} heterozygote. The average distance for all clones observed was 2.71 (± 0.73 , range 2–4, *n* = 14). c, *dsh* mutant clone induced in the presence of heat-shock promoter-*wg*. The average distance for all clones observed was 10 cells (1.6, range 8–12, *n* = 5). *hs-wg* produces insufficient activity to generate ectopic bristles in a wild-type background (not shown). The clones in b and c are not marked with *F^{66a}*, but are identified by the cell-autonomous tissue polarity defect caused by *dsh*¹⁴ (results not shown). Unmarked *dsh*^{ts} and *Notch*^{ts} *dsh*^{ts} clones were identified in



adult wings by the absence of normal margin bristles^{6,20}. 22/30 *dsh* clones were accompanied by ectopic bristles distant from the margin, as compared with 5/40 *Notch* *dsh* clones. The bristles near *Notch* *dsh* clones were only rarely as contiguously arrayed or as distant from the margin as those near *dsh* clones. Regulation within the margin proneural regions may account for these occasional bristles (data not shown).

FIG. 3 Interactions between *Notch* and *dsh*. a, b, *dsh* overexpression in the posterior compartment, using *UAS-dsh* and *en-GAL4*, caused narrowing and occasional loss of margin anti-Wg staining (b) when compared with anterior (a). *Dsh* levels, as identified using anti-Dsh, were visibly higher in the posterior compartment (not shown). Both panels are from the same imaginal disc. c, Model of Wg self-refinement. Notch signalling activity is highest at the dorsoventral boundary, and levels above the threshold (θ_{wg}) initially trigger broad *wg* expression (above). *Wg* represses *N* activity; *wg* expression is maintained only in cells nearest the dorsoventral boundary (below).



Guide to authors of contributions to *Nature*

Nature is a weekly international journal covering all the sciences. It is intended for an interdisciplinary readership, so all manuscripts should be written clearly and simply. Contributors should pay particular attention to the fact that English is not the first language of many readers. Space is limited and competition severe, so brevity is highly valued. The length limits below are specified in terms of the number of pages taken up in *Nature*. An uninterrupted page of text contains about 1,300 words, so a typical Letter to *Nature*, occupying 2½ pages, can contain about 1,500 words of text and four modest display items (figures and/or tables) with brief legends. There are no formal limits on the number of display items, but extra display material must be accompanied by commensurate reductions in the length of the text. Length limits can, however, be varied upwards or downwards at the editors' discretion.

Nature publishes the following types of contribution:

Review Articles survey recent developments in a topical area of scientific research. They should not generally occupy more than 6 pages of *Nature*.

Progress articles are relatively informal reviews of particularly topical and fast-moving fields. They do not exceed 4 pages of *Nature*.

Contributors wishing to submit a Review Article or Progress should first send a one-page synopsis to the Reviews Coordinator, from whom guidelines are available on request.

Articles are reports of original scientific research whose conclusions represent a substantial advance in understanding of an important problem and are of broad general interest. They do not generally exceed 5 pages of *Nature* and have no more than 50 references. They have a 'heading' of up to 80 words, often rewritten by *Nature's* editors, which advertises the content of the paper in general terms and should not contain numbers, abbreviations or measurements unless central to the message of the paper. The introduction and summary should be contained in the first three or four paragraphs of the main text. Layout of methods, figures and legends is summarized below.

Letters to *Nature* are short reports of original research focused on an outstanding finding whose implications are of interest to nonspecialists. They should not generally exceed 2½ pages of *Nature*. They have a bold first paragraph of not more than 180 words summarizing the background, rationale, conclusions and main implications of the work, aimed particularly at nonspecialist readers. Letters should have no more than 30 references. Layout of methods, figures and legends is summarized below.

Preparation of Articles and Letters

All manuscripts should be written in English, typed double-spaced on one side of the paper only. Manuscripts submitted in electronic form cannot be considered.

Manuscripts should be accompanied by a brief covering letter from the corresponding author, containing a full postal address, telephone and fax numbers, and e-mail address.

Five copies of manuscripts and original figures are needed, together with two copies of the covering letter. Five copies of manuscripts by any of the authors that have been submitted or are in press elsewhere should also be included, clearly marked as such; failure to do so may prejudice consideration of the manuscript.

Unless otherwise instructed by the editor handling a manuscript, when sending revised or resubmitted manuscripts five copies are required, each accompanied by a copy of the authors' response to referees' and editors' comments on the earlier version.

Titles should not exceed three lines of 31 characters each (including spaces), and should avoid numerical values, acronyms, abbreviations and punctuation wherever possible.

References should be numbered sequentially as they appear in the text, tables and figure legends. Only papers that have been published or submitted to a journal should be in the numbered list: manuscripts in preparation should be mentioned in the text with a list of authors. Reference lists should not contain textual material, grant details or acknowledgements.

Acknowledgements must be brief and appear after the reference list. *Nature* does not publish grant or contribution numbers.

Figures should not be larger than 22 by 28 cm unless unavoidable. They should not contain more than one panel unless the assembled parts are logically connected. Each copy should be marked with the corresponding author's name and, when known, the manuscript reference number. One photocopy of the original figures should be provided. All original figures are returned when *Nature* cannot offer to

publish a manuscript, but one copy of the manuscript and a photocopy of the figures are kept in confidential files for three months and then destroyed. Detailed instructions for figure preparation are available from the production department, including details of acceptable digital formats.

Colour figures are welcome but a contribution towards the cost of reproduction is requested. Inability to pay this charge will not prevent *Nature* publishing figures where colour is essential.

Tables. Symbols and abbreviations in a table should be defined immediately below the table, followed by essential descriptive material, all in double-spaced text. Tables should each be presented on a separate sheet of paper.

Methods. Brief details of materials and methods can be given in the text. Otherwise, they should be described at the end of the text in a section entitled "Methods", subdivided by bold headings referring to the methods used. This section should not exceed 800 words and ideally should be shorter. Additional material can be deposited on *Nature's* Web site. In contributions with methods sections, figure legends should not exceed 150 words each, and should be confined to a description of the figure and the symbols used.

Materials. As a condition of publication, authors are required to make materials and methods used freely available to academic researchers for their own use. Supporting datasets must be made available at the time of publication from the authors directly, by posting on *Nature's* Web site, by deposition in the appropriate public database or by distribution on the Internet. In the case of X-ray crystallographic coordinates only, public access may be delayed for up to one year after publication.

Articles and Letters should both contain a statement at the end of the text: "Correspondence and requests for materials to xxxx", with an e-mail address if desired, followed by the accession numbers or Web site address for supporting datasets. The designated author need not necessarily be the senior author, but *Nature* will expect the named author to handle matters arising from published contributions.

Manuscripts can be submitted to the Editor at *Nature*, Porter's South, Crinan St, London N1 9XW or at *Nature*, 968 National Press Building, Washington, DC 20045-1938. Proofs should be returned by express mail to London. Fax numbers and e-mail addresses for all editorial offices are on *Nature's* masthead.

Other contributions

Nature publishes informal contributed material as follows:

Correspondence contributions are short comments on topical issues, anecdotal material or reactions of readers to nontechnical material published in *Nature* (mainly News, Briefings, Opinion, Book Review or Commentary articles). All are unsolicited; submissions by e-mail to corres@nature.com are preferred (accents on names and places should be clearly indicated).

Commentary articles are informal, editorial comments on topical issues of public interest that have some bearing on scientific research. They are often controversial. Unsolicited manuscripts are considered, but enquiries can be made to the Commentary Editor by e-mail (nature@nature.com) or fax as a one-page synopsis or informal letter before formal submission.

News and Views articles inform nonspecialist readers about new scientific advances, either as reported in published papers or in the form of meeting reports. Most are commissioned, but proposals can be made to the News and Views Editor in advance of publication of the paper or of the meeting concerned. Authors are not allowed to discuss their own work or work from elsewhere in their institution. Detailed guidelines are available on request.

Scientific Correspondence is a predominantly peer-reviewed section of the journal specializing in very brief reports of general interest and discussion of topical issues. Priority is given to contributions of fewer than 500 words. Detailed guidelines are available on request. All contributions are unsolicited, and should be submitted by mail, not e-mail. Copies of contributions submitted by fax should be sent by mail with original figures and a cover letter noting that the manuscript has been faxed.

Book Reviews are all commissioned by the Book Review Editor. Unsolicited contributions are not considered.

Nature's editors will request final copy for Commentary, News and Views, Scientific Correspondence and Book Reviews on disk or via e-mail, as appropriate.

A more detailed Guide to Authors is available on *Nature's* WWW sites at <http://www.nature.com> or <http://www.america.nature.com>